Three Aspects of the Nature of Linguistic Meaning

by

Robert Kermode

B.Sc. Physics, McGill University, 1986;
M.A. Philosophy, McGill University, 1990


Submitted to the Department of Linguistics and
Philosophy in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy
in Philosophy

at the
Massachusetts Institute of Technology
February 1996

Signature of Author                                    . . . . . . .


Certified by  . . .                          . . . . . . .
                                                   Alex Byrne
            Professor, Department of Philosophy and Linguistics
                                              Thesis Supervisor


Accepted by . . . . . .        _           . . . . . . . . . . . .
                                                 George Boolos
                    Departmental Committee on Graduate Studies

Three Aspects of the Nature of Linguistic Meaning

by

Robert Kermode

Submitted to the Department of Linguistics and
Philosophy February 9th, 1996 in Partial Fulfillment of
the requirements for the Degree of
Doctor of Philosophy in Philosophy

ABSTRACT

This dissertation is in three chapters, each treating of a different aspect of the nature of linguistic meaning. Together, they support the thesis that linguistic meaning is a non-naturalistic property which we can have reason to ascribe only to the expressions of public languages.

The first chapter argues that meaning facts cannot be shown to be naturalistic facts by being shown to be reducible to teleological facts, and these in turn to naturalistic facts. Insofar as meaning facts can be reduced to teleological facts, it is because telelological facts are normative; insofar as teleological facts can be reduced to naturalistic facts, it is because they are non-normative. Fred Dretske, Ruth Garrett Millikan and Jerry Fodor are each criticized for missing this point.

The second chapter concerns rule-following and meaning scepticism. A sceptical argument based on epistemological considerations is set out and contrasted with Saul Kripke's argument for the conclusion that there is never a fact as to what one means by one's expressions. It is argued that the epistemological argument has broader implications than Kripke's. A solution to the sceptical problem raised, which relates one's epistemological access to the meanings of linguistic expressions to communal uses of the expressions, is presented in outline.

The third chapter develops the solution to the epistemological problem raised in the second chapter. The thesis that one's only epistemological access to the meanings of expressions is via communal use, is defended against two objections. The first objection is that, insofar as it can be shown that one cannot have reason to believe that one knows the meanings of the expressions of a language independently of knowing a community's uses of the expressions, it can also be shown that one cannot have reason to believe that one knows the meanings of a community's language. The second objection is that insofar as one can have reason to believe that one knows the meanings of the expressions of a language by knowing its communal uses, one can also have reason to believe that one knows them independently of any community. Each objection is responded to in turn.

Thesis Supervisor: Alex Byrne
         Title: Professor of Philosophy

ACKNOWLEDGEMENTS

My principal supervisor during the three years taken to write this thesis was Paul Horwich. Ned block also read and advised me on every draft. I am deeply indebted to them both for their guidance and instruction, as well as for their patience. In the last year of the writing of this, Alex Byrne kindly agreed to act as my third thesis committee member. I am much indebted to him for his help and advice, particularly on the third chapter. In addition to my thesis committee, I owe thanks for friendship and support to my colleagues, especially the members of my class and Darryl Jung and Emily Carson. In particular I am indebted to David Hunter, whose always careful and probing thinking made for many stimulating and illuminating discussions, and whose solid friendship helped me through MIT. I owe also, of course, an incalculable debt to my parents, who have supported me unquestioningly throughout my studies. By far my deepest debt is to Any Francois. Without her unfailing love and support, I would not have been able to finish this thesis. She has made the effort worthwhile.

My doctoral studies were generously funded by a fellowship from the Social Sciences and Humanities Research Council of Canada.

'

TABLE OF CONTENTS

# CHAPTER 1: TELEOLOGY AND MEANING NATURALISM

Common sense suggests that there are facts concerning what the words people utter, mean. When I say that the word 'pavement', as my friend Smith uses it, means *sidewalk*, I appear to be saying something that is either true or false, in the way that I would be if I said that Smith's weight is 160 pounds or that his car has a flat tire. According to one currently popular philosophical view, however, this common sense view can be true only if facts about meaning are, in a sense to be elucidated below, 'naturalistic' facts —if there are facts about meaning, then these facts must *ipso facto* be naturalistic. Since it is not obvious that meaning *is* naturalistic[1], it behoves anyone who subscribes both to the common sense view that there are facts about meaning and to the philosophical view that any such facts must be naturalistic, then, to show just how meaning is naturalistic. One popular strategy for trying to do this is to try to show that facts about meaning reduce to teleological facts of some kind, and then to try to show that these teleological facts reduce to naturalistic facts[2]. The aim of this chapter is to criticize this strategy —'teleological meaning naturalism', as we can call it— and in so doing, to cast doubt on the compatibility of the common sense view that there are facts

---

[1]Many reasonable philosophers have, of course, denied that meaning is naturalistic. See especially W.V. Quine, *Word and Object.* Cambridge, Mass.: MIT Press, 1960, and Saul Kripke, *Wittgenstein on Rules and Private Language.* Cambridge, Mass.: Harvard, 1982.

[2]In order to simplify the exposition, I will speak uncritically in what follows as though there are meaning facts and teleological facts. I do this solely for expository reasons, and do not mean in so doing to imply that I have answered the arguments of sceptics to the effect that there are no facts of these kinds.

about meaning and the philosophical view that any such facts would have to be naturalistic.

The motivation underlying teleological meaning naturalism is the thought that teleological facts are both naturalistic and normative, and thus will afford the means to naturalize what is perceived to be the normative aspect of meaning. As will be seen below in the course of reviewing three representative theories, however, some important proponents of teleological meaning naturalism entirely overlook the substantial problem of justifying this motivating thought. Instead, they take the central challenge facing them to be in demonstrating that meaning facts reduce to teleological facts. I will argue that insofar as any teleological fact is normative in the relevant sense, it is no more obviously naturalistic than are meaning facts. In short, I will argue that insofar as the problem of naturalizing meaning is in part the problem of naturalizing the normative, the strategy of appealing to teleological facts merely relocates the problem.

The first thing to do in what follows will be to elucidate the nature of the different kinds of (putative) facts in question —meaning facts, naturalistic facts, and teleological facts. It will then be possible to enquire as to whether biological science affords examples of appropriately normative teleological facts which are unproblematically naturalistic, which would clear the way for the teleological meaning naturalist. It will be found that it does not. Finally, three representative attempts to reduce meaning-facts to naturalistic facts by way of teleological facts, due to Fred Dretske, Ruth Garrett Millikan, and Jerry Fodor, respectively, will be criticised.

I

In order to be able to evaluate the claims of exponents of teleological meaning-naturalism, the terms of the discussion must be clarified, beginning with the kinds of putative facts for which a reductive analysis is being sought.

To see just what kind of facts about meaning are taken to be in need of reduction to naturalistic terms, consider Paul Grice's distinction between 'natural' and 'nonnatural' meaning[3]. Grice suggests that we distinguish between statements such as

"Those spots mean measles"

and

"Those clouds mean that it will rain",

on the one hand, and

"Those three rings on the bell mean that the bus is full"

and

"The meaning of the policeman's remark, "It is prohibited to
drive on the pavement", is that it is prohibited to drive on
the sidewalk",

on the other hand (Grice 1989, pp.213-14). As Grice notes, the concept of meaning which figures in the first statements, natural meaning, is such that $x$ (naturally) means that $p$ only if it is the case that $p$. For example, a certain cluster of dark clouds do not mean rain unless it actually will rain. To be sure, it may be that clouds of this type usually mean rain, but the very clouds in question —the 'token' clouds— do not mean rain in any sense, if it does not rain. In contrast, the concept of meaning which appears in

---

[3]Although Grice's terminology is, I think, apt, it should be clear that it is not meant to imply any conclusions as to the possibility of reducing facts about non-natural meaning to naturalistic facts.

the second statements, non-natural meaning, is such that a given
token of the sign type *x* may mean that *p* is the case, even if *p* is
not the case. For example, those token ringings of the bell may mean
that the bus is full even if a mistake has been made, and the bus
is not full. The first concept, natural meaning, can plausibly be
explicated in terms of the likelihood of *x* and *p* occurring together:
a given token of the type *x* naturally means that *p*, plausibly, if
and only if tokens of the type *x* are (considerably) more likely to
appear concomitantly with *p*'s being the case than not, and *p* is the
case. What is at issue in what follows, however, is the nature of
non-natural meaning. Just what sort of relation has to hold between
a sign type and a type of state of affairs, such that a token of the
sign type can mean that the state of affairs obtains, even when the
state of affairs does not obtain? Or, to put the point in terms
favoured by Fred Dretske, in what relation must a sign type stand
to the world, in order that tokens of the sign type can *misrepresent*
the state of the world[4]?

One suggested constraint on answers to the foregoing question is
that the relation at issue should, in some sense, be *normative*. That
this is the case is suggested by the following. Suppose that the
predicate 'green' in my idiolect means the property *green*. It does
not follow that all those things which I sincerely term 'green' will
be green —naturally, I may make mistakes. From the facts that I am
speaking sincerely and that 'green' in my idiolect means *green*,
then, we cannot conclude anything definite about what I will or will
not term 'green'. It appears, however, that we can draw specific
conclusions about what I should and should not term 'green' —about

---

[4]Cf. Dretske 1986, p.130.

what it would be *correct* for me to term 'green'. This is a point which Saul Kripke emphasizes in connection with the sceptical problem about meaning which he attributes to Wittgenstein[5]. In order for a theory of meaning to count as a candidate for solving this problem, Kripke maintains, it must be such as to entail, for any predicate 'P' of the language, conclusions as to what it would be correct to term 'P'. The so-called dispositional theory of meaning, for example, fails in this respect. This theory states, roughly, that for a word in a speaker's idiolect to mean something is just for the speaker to have a disposition to use it in certain ways. As Kripke argues, the dispositional theory fails inasmuch as it says nothing about whether it would be correct or incorrect for me to term any given thing 'green', whatever my linguistic dispositions.

A number of comments are in order regarding this point. The first comment concerns the sense in which normative consequences such as that one ought to term P's 'P' or that one has made a mistake in terming a non-P, 'P', *follow* from the fact that 'P' means P in one's idiolect. Note, to begin with, that there are possible situations in which one means P by 'P' but, all things considered, one ought *not* to term any P, 'P'. Such a situation might arise, for example, if terming a political dissenter 'a dissenter' would have the consequence of that dissenter's being executed. In such a situation, plausibly, even though 'dissenter' in one's idiolect means *dissenter*, it is not the case that one should term any dissenter 'a dissenter' —the norm that one should term P's 'P' is overridden by

---

[5]In his *Wittgenstein on Rules and Private Language* (Kripke 1982). See esp. pp. 24-32.

a stronger, moral norm. It seems, then, that normative consequences follow only 'ceteris paribus' from meaning facts. The sense in which the normative consequences follow is, in this respect, like the sense in which it follows from the fact that S has promised to do X, that S should do X. Here again, whereas this is ceteris paribus the case, there can clearly be circumstances in which the obligation is overridden. Now, it is not my purpose here to try to explicate this sense in which one fact can follow from another. I wish only to observe that, if it were to turn out that no explication of the relation can ultimatley be given, it would not follow straight away that there really is no such relation, or that it is somehow suspect or 'non-factual'. There are many properties and relations, such as goodness or knowledge, which, arguably, are insusceptible of a satisfactory philosophical analysis, but which we should nevertheless not, merely on this account, take to be suspect.

The second comment concerns whether a meaning naturalist might not simply circumvent the difficulty of naturalizing the normative aspect of meaning altogether. The naturalist might argue, that is, that the appearance of the normativity of meaning can be accounted for by the fact that speakers adhere to the very basic principle that, ceteris paribus, one should speak truthfully, and hence need not be due to any normativity inherent in meaning. Now, whatever the promise of this strategy generally, it is not of interest to the teleological meaning naturalist. This is because the plausibility of teleological meaning naturalism lies precisely in its (supposed) ability to accomodate the normative aspect of meaning. That is, the teleological meaning naturalist's claim is that what *makes it plausible* that tokens of a given sign type 'S' in organism O's

idiolect[6] mean that it is the case that $W_r$ is that, given O's 'design', O *should* produce a token 'S' when it is the case that W (or, O *makes a mistake* if it produces a token of 'S' when it is not the case that W). Roughly speaking, inasmuch as the (supposed) fact that O and the sign type 'S' have certain teleological properties entails (ceteris paribus) that O should produce a token of 'S' when it is the case that W, and inasmuch as 'S''s meaning that W is the case in O's idiolect also entails this, it is taken to be plausible that for O and 'S' to have these teleological properties is just for 'S' to mean that W is the case in O's idiolect. Teleological meaning naturalists, then, have no interest in denying that facts about meaning entail (ceteris paribus) normative implications.

The third and last comment concerns the nature of the normativity in question. One might, with some plausibility, maintain that the sense in which one should (ceteris paribus) term only P's 'P' if one means P by 'P', and the sense in which one should, say, avoid harming unnecessarily one's neighbours, are different senses. In the former case, one might maintain, one should do the action in question because it is of instrumental value; in the latter case, because it is one's moral duty. Roughly speaking, the difference is between the presumption in favour of doing what is practical or expedient, and the presumption in favour of doing what is *right*. And, one might maintain that whereas the latter type of normativity is problematic for the naturalist, the former type is not. Thus, according to this line of thought, the view that naturalizing meaning is made problematic by the need to naturalize the normative derives from a confusion as to the nature of the normativity in

---

[6]Or langauge of thought, as the case may be.

question. Now, the primary difficulty with this line of thought is not with the distinction between 'instrumental' norms and 'pure' norms[7]. The difficulty is, rather, with the claim that there is no problem in naturalizing the 'instrumental' normative consequences which follow from meaning facts. For, such instrumental normative claims do presuppose that the agent subject to the norms has interests or goals. For example, insofar as it is the case that an agent should avail herself of a ¼" wrench, it is because the agent has some goal, such as to loosen a ¼" bolt. And it is these mundane interests or goals which, I shall maintain, are insusceptible of naturalistic reduction.

The basic problem, again, is to show that facts about what people mean reduce to uncontroversially naturalistic facts. It is reasonable to ask, before proceding further, just what 'naturalistic facts', are. Formulating a general answer to this question, however, is a substantial project in itself. Fortunately, since the aim here is merely to show that teleological meaning naturalism is an unworkable strategy, it will be sufficient to give a general characterization which includes all those kinds of facts which teleological meaning naturalists themselves take to be uncontroversially naturalistic[8].

Although meaning naturalists are not typically very voluble on the subject, there appears to be agreement that at least the following two kinds of facts count as naturalistic. First, there are

---

[7] I leave open the question as to whether this distinction can ultimately be defended.

[8] By 'uncontroversially naturalistic' I mean naturalistic and not in need of being shown to be such. Teleological meaning naturalists take meaning facts to be naturalistic, but grant that they are not uncontroversially so.

facts concerning the dispositions of things. As Nelson Goodman indicates, there are a wide variety of facts of this kind:

> Perhaps we ought to notice at the very beginning that more predicates than we sometimes suppose are dispositional . . . To say that a thing is hard, quite as much as to say that it is flexible, is to make a statement about potentiality. If a flexible object is one capable of bending under pressure, a hard object is one capable of resisting abrasion. And for that matter, a red object is likewise one capable of certain color appearances under certain lights; and a cubical object is one capable of fitting try squares and measuring instruments in certain ways. (Goodman 1983, p.40)

Other examples are that something has a given mass or charge, that it exerts a certain gravitational force, or simply that it is fragile or slippery. Inasmuch as each of these facts concerns the "potentialities", or "threats and promises" which a given thing holds, they are all what might be called dispositional facts. What all of these facts have in common, is that in each case, that the object or substance in question has the property in question is directly ascertainable by testing or measuring the object or substance. Now, one way of cashing out what it is for something to have a disposition is to specify what would be true of it if conditions had been otherwise. For example, for something to be flexible at time $t$ is for it to be such that, had it been subjected to suitable pressure at $t$ it would have bent[9]. This suggests that we should also include, together with dispositional facts, facts about what would be the case if conditions were otherwise —i.e., the facts stated by counterfactual conditionals.

The second class of facts which are uncontroversially naturalistic, etiological facts, are unlike the first kind in respect of their testability. By 'etiological facts', I mean such

---

[9]Goodman 1983, p.35 —though see also Goodman's qualifications on this point, pp.36-9.

facts as that marsupials evolved in Australasia, that certain
mountains were formed by volcanic forces, or that a certain painting
is an original Matisse. In each of these cases, what is in question
is the causal history of the object. That a fact such as this
obtains, evidently, is not ascertainable merely by testing the
object(s) in question. Tests may often provide very good evidence,
but will underdetermine whether the fact in question obtains to the
extent that other causes can give rise to similar objects.
Nevertheless, although there are perhaps contexts in which the
factuality of etiological claims might be questioned[10], they are
unproblematically factual from the point of view of the naturalist.

An example of a class of claims which are *not* uncontroversially
naturalistic is the class of normative claims. Examples of such
claims are that one ought not to kill, that one ought to keep one's
promises, and that one should dress warmly in cold weather.
Arguments which purport to show that such claims do not state
naturalistic facts, or do not state any facts at all, are
familiar[11]. Arguments such as these, I take it, have been
considered to be sufficiently persuasive that it is considered
incumbent upon anyone who maintains that normative claims state
naturalistic facts, to *show* that normative facts reduce to
uncontroversially naturalistic facts.

Now, if teleological meaning naturalism is correct, then
teleological facts must reduce to uncontroversially naturalistic

---

[10]I.e., where scepticism about the past is an issue.

[11]See, for example, Moore's argument in *Principia Ethica*.
Cambridge: Cambridge University Press, 1903, esp. §§5-14, and J.L.
Mackie's argument in his *Ethics: Inventing Right and Wrong.* New
York: Penguin, 1977, esp. Ch. 1.

facts such as of the two kinds just cited. To see whether they do, it is important first to get clear on the conception of teleological facts at issue. When teleological meaning naturalists speak, say, of an organism's brain's going into a certain state as having a certain function or purpose, they evidently do not mean that the organism's brain's going into the state is part of some grand design for the Universe —that it serves some interest beyond the organism's (or its genes' or species') own. Putative facts about the Universe's design would evidently be less likely to be naturalistic than are the meaning facts they are supposed to explicate. Similarly (perhaps equivalently), they clearly do not mean that the organism's brain's going into the given state is somehow "caused" by some such thing as a destined end state of the Universe —that it is the effect of some "final cause". That is, they evidently do not intend to imply that there are causes at work beyond those countenanced by contemporary natural science. What teleological meaning naturalists apparently *do* mean in saying that an organism's brain's going into a certain state has a certain function, is rather that it (the organism's brain's going into the state) subserves the achievement of some purpose or aim or goal, or the fulfillment of some need, of the organism in question. The functional attributions in question, then, are much like, for example, the attribution to the metal fins on the back of an audio amplifier of the function of dissipating excess heat: serving the function of dissipating excess heat, and thereby keeping the amplifier at a suitable operating temperature, conduces to the fulfillment of the amplifier's purpose of magnifying with minimal distortion an input signal.

There are two points to be remarked here. First, as has already been suggested, teleological facts appear to have normative consequences. Thus, for example, from the facts

(i) that a system *S* has as its purpose to attain some end state *E*,

and

(ii) that *S* behaves on an occasion in such a way as to make the attainment of *E* unlikely,

it appears to follow directly that

(iii) *S* behaves *improperly* —that *S* has *failed* in its purpose.

Now, as with the supposed normativity of meaning, one could argue that the normative nature of this conclusion arises not from any inherent normative character of teleology, but from a suppressed premise to the effect that one ought to act so as to fulfill one's purposes[12]. Insofar as the teleological facts in question are taken to serve to naturalize the normative aspect of meaning without any such additional premise being assumed, however, this view is not open to teleological meaning naturalists.

The second point is that it is essential to distinguish between something's having a teleological property, and its merely *appearing* to have a teleological property. That something should appear to have a purpose by virtue of its behaviour, does not entail that it does. To illustrate this point, consider a machine which has two tracks, tank-style, which propel it, and two light sensors in place

---

[12]That is, one could argue that the principle, 'Ceteris Paribus, (*x*)(*p*)[(*x* has purpose *p* and *x* acts so as to undermine the achievement of *p*) → (*x* acts improperly)]', if true, is so only in virtue of some general normative premise(s), such as 'One ought, other things equal, to act so as to fulfill one's purposes'.

of eyes[13]. The light sensors are cross-connected to the track motors, in such a way that light shining on the left sensor increases the speed of the right track, and light on the right sensor increases the speed of the left track. The result of this is that the machine tends to steer toward light sources. When set in motion in the presence of a light source, then, this machine will have the appearance of having as its aim or goal to reach the light. This appearance, however, is compatible with the machine's having quite a different purpose, as would intuitively be the case if it had been designed to avoid light but had been miswired, or no purpose at all, as would intuitively be the case if the machine somehow came into existence by purely random processes. This point applies, furthermore, to systems which exhibit behaviour much more complex than that of the machine just described. It seems entirely possible that a system should behave in quite complex ways which appear to be goal-seeking, and yet have no goal at all. Indeed, sceptics about teleology maintain that *no* system, no matter how complex the behaviour exhibited, has a purpose in any of the senses just considered[14].

Proponents of teleological meaning naturalism might well object to the foregoing that they never supposed that the types of teleological facts just discussed are clearly naturalistic. Indeed, they might well concede that they are not naturalistic. What they had in mind, they might say, are facts about so-called *biological*

---

[13]This example is drawn from from Valentino Braitenberg's book *Vehicles* (Cambridge, Mass.: MIT press, 1984), which describes a series of machines which are meant to illuminate the evolution of certain common biological phenomena.

[14]Cf. W.V. Quine, *Pursuit of Truth*. Revised ed'n. Cambridge, Mass.: Harvard, 1992. p.75.

functions or purposes or needs, which *are* naturalistic. Unlike the
teleological facts discussed above, the story goes, biological
functions or needs do not make the problematic supposition that the
organisms in question have purposes —all they suppose is that the
function-possessing subsystems (organs) of organisms contribute to
the organisms' reproductive success. The problem which this raises
for the meaning naturalist, however, is that it appears that such
deflated functions are not capable of naturalizing the normative
aspect of meaning[15]. The criticism being levelled at teleological
meaning naturalists in this chapter is, in effect, that they
equivocate between the two conceptions of the functional facts in
question. When defending the naturalistic credentials of their
reduction, they take the facts to be teleological only in this
deflationary, 'biological' sense. The facts they invoke in reducing
the normative aspect of meaning, on the other hand, are, of
necessity, teleological in the inflationary, normative sense.

In the next section, it will be argued that the notions of
function or purpose which figure in biological science do not
presuppose that its possessors are subject to norms. It will then
remain to show that teleological meaning naturalists really do need
a more 'inflated' notion of function to effect their reductions.


## II

One obvious reason for thinking that at least some teleological
facts are naturalistic, is that i)statements about the functions of

---

[15]Again, the target of this paper is the theorist who grants
that meaning is normative. It is possible that one might attempt to
naturalize semantic notions non-normatively conceived, by means of
biological functions —I reserve judgement on such a strategy here.

things appear to be teleological statements, ii)such statements sometimes figure in biological explanations, and iii)biological explanations are evidently naturalistic. As Robert Cummins (Cummins 1975) has convincingly argued, however, the reasons for thinking that the notion of function appealed to in biological explanations is a genuinely teleological notion, are bad reasons, and there are alternative accounts of the notion —one of which Cummins gives— which are *not* teleological[16]. Although it should not be surprising, on reflection, that biological science does not entail that its objects of study are subject to norms or have ends, it is important to see, on the one hand, how the appearance can arise that it does, and on the other hand, how statements about the functions of things can appear in biological explanations without the explanations' *ipso facto* being teleological.

An illustrative example of a statement concerning the function of a thing, such as might be appealed to by a biologist, is,

(1) The function of the lungs in organisms with lungs is to oxygenate the organisms' blood.

What is in question is whether this biological claim presupposes that the organisms of which the lungs are subsystems have some end or ideal state, the attainment of which the lungs contribute toward, and which would serve to ground the normativity of meaning. Cummins identifies two assumptions common to analyses of function-ascribing statements which can suggest that this is the case[17]. These are,

------

[16]To avoid confusion, in what follows, only those facts concerning functions which have the normative consequences discussed above, will be called 'teleological'.

[17]Cummins argues that these two assumptions are false. As my concern is merely to show how they can be misleading, I will not attempt to defend Cummins' criticisms.

(A) The point of functional characterization in science is to explain the presence of the item (organ, mechanism, process or whatever) that is functionally characterized.

and

(B) For something to perform its function is for it to have certain effects on a containing system, which effects contribute to the performance of some activity of, or the maintenance of some condition in, that containing system. (Cummins 1975, p.741)

Consider first assumption (A). Where statements such as statement (1) are concerned, the plausibility of (A) is apparent: the anatomist, upon discovering the presence of lungs in a wide variety of animals, will naturally want to know why they occur where they do. An account of the function or purpose of the lungs in the animals' overall physiology seems to afford an answer —they are present because they fulfill the function of oxygenating the blood. It is a familiar point, though, that the 'because' here cannot be the 'because' which figures in straight-forward deductive-nomological explanations, inasmuch as something other than lungs might have fulfilled this function[18]. Some other type of explanation must be in play here, then, supposing that the lungs' function does somehow explain their presence. One tempting alternative is to construe the explanation on the model of explanations of the presence of components of artefacts produced by rational agents, and it is this alternative construal which is one source of the idea that the notion of function appealed to in science is teleological. That such a construal cannot be countenanced by biological science becomes apparent when this explanatory model is examined.

---

[18]More specifically, the objection is that from the fact that *x* has as its function to *y* and the fact that function *y* is fulfilled in system *S*, it does not follow that *x* is present in *S*.

Where artefacts are concerned, the point of explanations of the presence of their components is characteristically to supply the artefact's creator's *reason* for putting the component in the artefact. Where an explanation of some component's presence appeals to the component's *function*, this is characteristically because the artefact was made to do something —to achieve or help achieve some end— and to cite the component's function is to say how the component contributes to the artefact's achieving its end, and hence to give the creator's reason for putting it there (Cf. Cummins 1975, p.746). For example, if I ask, concerning the carburetor in a car engine, 'Why is this here?', the answer will be that it is there (was put there by the designer) to fulfill the function of mixing air with the gasoline, and hence to permit combustion and the running of the engine. Where explanations concerning the presence of components of natural systems are concerned, in contrast, there is no designer, the behaviour of whom to rationalize, so this kind of explanation is inappropriate. To be sure, it is sometimes said that 'Mother Nature' designed organisms with lungs to provide for the oxygenation of their blood, but this is at best a loose metaphor for the process of evolution. If the aim is to naturalize meaning facts via teleological facts such as facts about the functions of things, it evidently will not be acceptable to explicate the latter merely in terms of 'Mother Nature's intentions'.

If appeals to 'Mother Nature's design' have any substance, it is clearly insofar as they are just a colourful way of talking about the effects of evolution. To get clear on the relation between evolution and function, consider next assumption (B). *Prima facie*, it is plausible that something like this assumption explains the

role of functional explanations in biology, and that it does this in such a way as to entail that the notion of function used in biology is teleological. Careful consideration, however, shows that interpretations of (B) which construe functions as teleological either fail to specify uniquely the functions in question, or else make unjustified assumptions. This becomes apparent in attempting to elucidate (B). The natural general analysis of function ascribing statements suggested by (B) is,

> (2) The function of an $F$ in a $G$ is $f$ just in case (the capacity for) $f$ is an effect of an $F$ incorporated in a $G$ in the usual way ... and that effect contributes to the performance of a function of the containing $G$. (Cummins 1975, p.753)

Applying this to lungs, we get

> (3) The function of the lungs in an organism is to oxygenate the organism's blood just in case oxygenating the blood is an effect of lungs incorporated in an organism in the usual way, and oxygenating the blood contributes to the performance of a function of the organism.

This looks like a promising beginning, except that it makes the truth of (1) contingent upon whether there is some function of the organism, the performance of which is contributed to by the oxygenation of the blood. As things stand, then, (2) yields an analysis of the function of the lungs in terms of some unspecified function of the organism as a whole, and hence merely shifts the problem of determining the nature of the functions in question. There are two options for dealing with this.

The first suggestion for rectifying (2) is to modify it by dropping the reference to the function of the containing system:

> (2') The function of an $F$ in a $G$ is $f$ just in case (the capacity for) $f$ is an effect of an $F$ incorporated in a $G$ in the usual way.

The problem with this suggestion is that $F$'s will invariably have effects which are not in any intuitive sense their function. An example where lungs are concerned is the small air movements in the region of the nostrils which they produce. According to (2'), it is part of the function of the lungs to produce the small air movements which they happen to produce, and this is intuitively not plausible. Indeed, the point of specifying that the effect of $F$ which is its function is that effect which contributes to some function of the containing $G$, was precisely to rule-out such incidental effects as the small air movements caused by the lungs. (2'), then, is not the solution. The second suggestion for rectifying (2) is somehow to specify the function of the containing system. A suggestion along these lines for biological organisms which Cummins considers, and which surely underlies the thinking of many people on the subject, is that the relevant function of an organism is to sustain or increase its species' numbers. This suggests

> (2'') The function of an $F$ in a $G$ is $f$ just in case (the capacity for) $f$ is an effect of an $F$ incorporated in a $G$ in the usual way and $f$ contributes to the organism's ability to sustain or augment its species' numbers (Cf. Cummins 1975, p.755).

Since the lungs oxygenating an organism's blood contributes to the organism's staying alive and healthy, which in turn enhances its ability to augment its species numbers, (2'') yields (1), as desired. (2''), then, appears to be an acceptable explication of assumption (B).

Despite its apparent acceptability, (2'') is potentially misleading. One might try to defend (2'') on the grounds that it is extensionally correct —i.e., that those capacities of subsystems of organisms which (2'') specifies as their functions, are indeed the

capacities which biologists claim are their functions. One problem with this justification, as Cummins observes, is that it fails entirely to tell us why *f* is taken to be the *function* of *F*. As he says, concerning what amounts to this suggestion,

> ...[it] would at most tell us *which* effects are picked out as functions; it would provide no hint as to *why* these effects are picked out *as functions*. We know why evolutionary biologists are interested in effects contributing to an organism's capacity to maintain its species, but why call them functions? (Cummins 1975, p.750).

Now, as far as the present discussion is concerned, if one were to balk at this criticism, and assert that (2″) is the correct analysis simply in virtue of its extensional adequacy, there need not be any objection. By itself, (2″) is compatible with the supposition that organisms have no aims or purposes and that their subsystems' having functions entails no normative consequences, which is all that is being claimed here. Insofar as (2″) is taken to indicate *why* the relevant capacities of subsystems of organisms are taken to be *functions*, however, it depends on an unjustified teleological assumption —viz., that the function or aim or purpose of an organism is to sustain or increase its species' numbers. It is this assumption which yields (2″) from (2). Evidently, however, this assumption is not justified by evolutionary theory: evolutionary theory makes no assumptions and draws no conclusions about aims or purposes whatsoever. What the theory of evolution says is, to a first approximation, no more than that if the propensity of lungs to oxygenate blood is more conducive to survival than not, then, among creatures otherwise identical, those with lungs will proliferate at the expense of those without. Any assumption to the effect that it is an organism's aim or purpose to survive is evidently extraneous to biology.

This point, it should be noted, is not affected if one takes natural selection to act solely or primarily at the level of genes, rather than at the level of organisms and species[19]. Genes, evidently, no more have the sorts of aims or purposes or goals which would fix the functions of, say, organs or cell structures or proteins, and thereby determine norms for their behaviour, than do organisms. To imbue complex proteins jostled about in cell nucleii with purposes, and to regard, say, a site on a given replicated DNA molecule which does not match the corresponding site on the molecule from which it was replicated, as a *mistake*, is clearly no more than to adopt a *façon de parler*, albeit one possibly with considerble heuristic value.

Given the difficulties with the suggestions reviewed so far, Cummins recommends rejecting (A) and (B) in favour of an alternative account of functions as they occur in scientific explanations. Cummins' suggestion is that the purpose of functional explanations in natural science is to explain given capacities of containing systems, relative to analyses of the containing systems into subsystems. More specifically, a functional explanation explains how it is that a given containing system has a given capacity, by appealing to the capacities of its component subsystems, where the capacities of the subsystems appealed to are taken to be their functions relative to the analysis. In Cummins' words

> [subsystem] $x$ functions as a $\Phi$ in [containing system] $s$ (or: the function of $x$ in $s$ is to $\Phi$) relative to an analytical account $A$ of $s$'s capacity to $\Theta$ just in case $x$ is capable of $\Phi$-ing in $s$ and $A$ appropriately and adequately accounts for $s$'s capacity to $\Theta$ by, in part, appealing to the capacity of $x$ to $\Phi$ in $s$. (Cummins 1975, p.762)

---

[19]For an exposition of this view, see, for example, Richard Dawkins *The Selfish Gene*. Oxford: Oxford University Press, 1976.

If this conception of functional explanation in science is correct, then what biological function a bodily organ has is relative to biologists' explanatory purposes, and not to any putative purposes of the organism or organ in question. For the present, what is interesting about this account is that it in no way entails that the system in question has a purpose or end or goal. Insofar as a non-teleological account of biological functions is to be preferred to a teleological one, given concerns about whether teleological facts are naturalistic, this account is *prima facie* preferrable to any teleological account. Any proponent of the view that biological functions are teleological, then, has at least the burden of showing this account of Cummins' to be false.

## III

In the remainder of this chapter, I propose to consider three representative examples of teleological meaning naturalism. I will argue that none of these makes a convincing case that the teleological facts required are naturalistic, and hence that each fails.

The first example of teleological meaning naturalism to be considered is due to Fred Dretske. In his paper 'Misrepresentation' and in his *Explaining Behavior*, Dretske sets out a theory of what it is for something —a brain state of some animal, say— to *mean* that the world is a certain way. The theory is intended to be a naturalistic theory which solves certain underdetermination problems. As will be seen, though, for present purposes, what is interesting about Dretske's exposition is not primarily his theory but rather his conception of the problem which the theory is

supposed to solve: he takes the problem to be in reducing meaning facts to teleological facts, and does not address the problem of reducing teleological facts to uncontroversially naturalistic facts.

The basic task which Dretske sets is to explain how it is possible "for physical systems to *misrepresent* the state of their surroundings" (Dretske 1986, p.129). Before setting out and evaluating Dretske's proposal, two comments are in order. First, Dretske focusses on *mis*representation, rather than representation, because it is that relation which holds between a sign token and a state of affairs even though the state of affairs may not obtain, which is at issue. As was observed in section I above, the notion of meaning at issue here is that according to which a token of a sign type can mean that a certain state of affairs obtains even when the state of affairs does not obtain —i.e., can misrepresent the state of the world (Dretske 1988, pp.64, 5).

The second comment is that, as Dretske observes, it is necessary to distinguish here between the representational powers of such things as maps and signposts, which are parasitic upon the representational powers of intentional agents, and the underived or 'intrinsic' representational powers of a certain privileged set of signs (Dretske 1988, pp.64, 5). Where maps and signposts are concerned, their representational powers are evidently parasitic on those of intentional agents, inasmuch as it is only because intentional agents *assign* them the function of representing something, that they are anything more than mere colourful bits of paper or wood or whatever. Assuming, however, that not all representational powers can be parasitic in this way, on pain of infinite regress, there must be some signs whose representational

powers are intrinsic. The problem at hand is to determine how it is that physical systems can have this intrinsic power to represent and misrepresent the world. As Dretske says, "[w]hat we are after is *nature's* way of making a mistake, the place where the misrepresentational buck stops." (Dretske 1986, p.131).

The first step in Dretske's account is to draw a distinction between what he calls "natural" and "functional" meaning —henceforth 'meaning$_n$' and 'meaning$_f$', respectively. Dretske's notion of natural meaning is, as he acknowledges, just the one already encountered, due to Grice: a token of a natural sign can mean$_n$ that some state of affairs obtains, only if the state of affairs actually obtains. Thus dark clouds mean$_n$ that it will rain, only if it really will rain. Insofar as meaning$_n$ is a relation which does not hold between a token natural sign and some state of affairs if the state of affairs does not obtain, it will not suffice to give an account of misrepresentation. Thus Dretske introduces meaning$_f$, which he defines as follows:

> *d's* being *G* means$_f$ that *w* is *F* = *d's* function is to indicate the condition of *w*, and the way it performs this function is, in part, by indicating that *w* is *F* by its (*d's*) being *G*. (Dretske 1986, p.133)

For example, the alcohol in a thermometer being at the height of the "0 °C" mark means$_f$ that the ambient temperature is 0 °C, provided that it is the function of the thermometer to indicate the ambient temperature, and the thermometer does this, in part, by having its alcohol at the height of the "0 °C" mark when the temperature is 0 °C. The crucial point here is that it remains the case that the thermometer's alcohol's being at this height means$_f$ that it is 0 °C, even if the thermometer is somehow defective and this happens when the temperature is not 0 °C, provided that the thermometer continues

to have as its function to indicate the temperature. In short, the thermometer can misrepresent the state of the world.

The problem posed by the definition of meaning$_f$, as Dretske acknowledges, is to give a naturalistic account of what it is for *d* to have the function of indicating the condition of *w*. To illustrate what he takes to be the difficulty in solving this problem, Dretske proposes, and then criticizes, a relatively simple trial theory[20]. His trial suggestion is that, in some cases at least, the function of some subsystem (organ) of an organism is to fulfill that information-gathering role which it has fulfilled in the course of the evolution of the organism, whereby the organism has been able to satisfy some 'biological need'. A biological need, here, is, "some thing or condition without which [the organism] could not survive" (Dretske 1986, p.135). To illustrate how this theory is supposed to work, Dretske gives an example.

The example involves a certain type of marine bacteria found in the Northern Hemisphere. These bacteria have organelles called 'magnetosomes', which are magnetized and hence line-up with the prevalent (geo)magnetic lines of force. Provided that the bacteria are in the Northern hemisphere, the effect of this behaviour of their magnetosomes is to orient the bacterial cells such that they propel themselves away from the oxygen-rich water surface and toward the oxygen-poor depths. Since these bacteria cannot survive in the presence of oxygen, this is conducive to their survival. As Dretske observes, it may be tempting here, given the suggested theory, to infer that it is the *function* of the magnetosomes to indicate the

---

[20]Dretske explicitly compares this trial theory to the theory of Fodor's which will be discussed below (Dretske 1986, p.136).

direction of relatively oxygen-poor water. According to this suggestion, then, a magnetosome's (d's) being aligned in a given direction (being G) means, that the water (w) is relatively oxygen-poor in the given direction (is F), inasmuch it would be the function of the magnetosome's alignment to indicate the direction of oxygen-poor water. Thus, if the magnetosome should be aligned such that the water is *not* oxygen-poor ahead (as would happen if the bacterium were displaced to the Southern hemisphere), then the magnetosome misrepresents the environment to the bacterial cell in which it resides.

It has already been seen, in section II above, that the theory of what it is for a subsystem to have a function illustrated by this example, is inadequate to the naturalist's purposes. On the one hand, suppose that the organism's 'needs', and hence also the organism's subsystem's functions, are taken to be genuinely teleological —i.e., that they presuppose that the organism has some end which it is aiming to fulfill, such that if the organism or its subsystem behaves so as to undermine the fulfillment of the end, it errs. In the present case, to suppose this is to suppose that the bacteria in question have it as their aim to avoid oxygen rich waters. It must be asked, supposing this, just which naturalistic facts entail that the bacteria have the *aim* of avoiding oxygen rich waters. Dretske says, without further explanation, that an organism has a need of something if having the thing is a precondition for its survival. Evidently, however, this is true, given the suggested sense of 'need', only if some *value* is placed on survival —if survival is a matter of indifference, then the preconditions for survival can hardly be called 'needs' in the given sense. As long

as an organism is regarded merely as a complex organic system, capable of being in various states of which some correspond to its being alive, others dead, as biological science recommends, there is no question of its having aims or needs in the suggested sense. In short, the criticisms which were made in section II of proposal (2″) apply here —the account goes through only if ultimately, the natural but not naturalistic assumption is made that organisms have it as their *aim* to survive. On the other hand, suppose that the needs and functions here are not teleological in the relevant sense. The problem in this case will be that the grounds for taking *d*'s being *G* when *w* is not *F* to be a *misrepresentation*, will be lost. If it is not in any sense the aim or purpose of magnetosomes to indicate the direction of oxygen-poor water (or whatever), then a given magnetosome's being oriented away from the oxygen-poor water is at most merely an occurence of an improbable (although statistically predictable) event. Intuitively, something more is needed to make it a misrepresentation —a misrepresentation is a *mistake*, a failure to represent properly[21]. Dretske's terminology at the beginning of 'Misrepresentation' reflects this thought:

> "Epistemology is concerned with knowledge: how do we manage to get things right? There is a deeper question: how do we manage to get things wrong? How is it possible for physical systems to *misrepresent* the state of their surroundings?" (Dretske 1986, p.129).

To misrepresent the state of one's environment is to get things wrong. Getting things wrong, though, requires having interests or values, and that a bacterium or human has these is not implied by biology.

---

[21]In the course of evaluating Fodor's theory, below, the suggestion will be evaluated that, notwithstanding this intuition, nothing more is needed for representation.

The problem which Dretske finds with his trial theory is not the problem just rehearsed. The difficulty which he finds, is that there is no principled reason for choosing, as the function of the magnetosomes, to indicate the direction of oxygen-poor water. Why not, Dretske says, take it to be to indicate the direction of geomagnetic north, or even just the direction of north in the local magnetic field (which might be caused by a bar magnet)? Given the environment in which the bacteria evolved, the magnetosomes' having been indicators of any of these things would have conduced to the proliferation of its kind. But, Dretske observes, which function the magnetosome has affects crucially what it represents. Suppose that a bacterium is displaced to the Southern Hemisphere. If the function of its magnetosome is to indicate the direction of oxygen-poor water, then the magnetosome misrepresents the bacterium's environment when it points north, whereas if its function is to indicate the direction of geomagnetic north, it correctly represents the bacterium's environment. According to Dretske, the upshot of this is that the basic evolutionary account of representation leaves indeterminate what the sign in question means$_f$.

Insofar as Dretske concludes that the basic evolutionary properties just considered of an organ or subsystem of an organism, do not entail that it has any particular function, his position appears to be consistent with the thesis of this chapter, that teleological properties are not reducible to naturalistic properties. This, however, is not Dretske's ultimate view. Although he does maintain that where certain simple organisms, such as the bacteria just considered, are concerned, the functions of their organs is indeterminate, he argues that where more complex organisms

are concerned, there are naturalistic properties in addition to the evolutionary properties already discussed which fix the functions of the organisms' internal organs or subsystems. The relevant question here is whether Dretske, in trying to solve the indeterminacy problem, also succeeds in answering the kind of objection raised just above.

Dretske's proposal is that what is needed in order that the function of an organ or subsystem should be fixed, is first, that there be redundant mechanisms by which the subsystem's being in a certain state indicates the relevant state of the world, and second, that the indication relation which holds between the given state of the subsystem and the world, be acquired as the result of associative learning. The first of these requirements, more specifically, is that for $d$ to have as its function to indicate the state of $w$, there must be more than one 'route' by which the state of $d$ indicates the state of $w$. Adapting slightly Dretske's illustrative example of this, imagine an organism which has evolved to avoid a certain kind of plant. The organism has a brain mechanism which goes into a certain state *either* when the organism's sensory organs are stimulated by (what smells like) the plant's buds in the springtime *or* (what looks like) the plant's leaves in the summertime. Since the mechanism's being in the given state is a reliable indicator of the presence of the plant but not either of the buds or of the leaves alone (since, depending on the season, either the leaves or the buds will be absent), Dretske concludes —with one qualification— that it is the function of the state of the mechanism to indicate the presence or absense of the plant. The qualification, which is the second part of Dretske's proposal, is

required to preclude the possibility that the function of the mechanism's being in the given state is to indicate the disjunctive property of having a certain smell *or* looking a certain way. The additional requirement, that the state must come to indicate the relevant state of the world through associative learning, serves to select as the state represented, the common cause of the avoidance behaviour, of which the smell and look are mere symptoms.

The question at hand, again, is whether Dretske convincingly demonstrates that it is the *function* of the given organ or subsystem to indicate the given state of the world, if the two stand in the appropriate naturalistic relation to one another. I do not see that he does. Of the naturalistic properties appealed to, I do not see that any entails that either the organ or subsystem, or the containing system (organism), has any end, goal, function or purpose at all. That a subsystem *d* comes, as a result of associative learning over time, to go into state *G* when *w* is *F*, in no way entails that it is the purpose of *d* to be in *G* when *w* is *F* —*d* might well have no purpose at all. Nor do matters change if *d* indicates the state of *w* via more than one informational route, as in the example of the plant-avoiding organism above. That *w*'s being *F* should induce *d* to become *G* via more than one causal pathway does not in any obvious way entail that it is *d*'s purpose to indicate the state of *w*.

The problem with Dretske's account is evident. After having rejected the basic trial theory of representation which he initially considers, Dretske construes the problem thus:

> "...the sticky question is: *given* that a system needs *F*, and *given* that mechanism *M* enables the organism to detect, identify cr recognize *F*, *how* does the mechanism carry out this function? Does it do so by representing *nearby Fs as nearby*

*F*s or does it, perhaps, represent them merely *as nearby G*s,
trusting to nature (the correlation between *F* and *G*) for the
satisfaction of its needs?" (Dretske 1986, p.139. Dretske's
emphases)

A stickier question than this, I have been arguing, is whether facts
about a thing's needs are naturalistic facts, and hence can be
accepted within the context of a naturalistic theory —whether it can
be "given" that a system needs *F*. If such facts are naturalistic,
as Dretske assumes, then, plausibly, so also are facts about the
functions of organisms' subsystems, since, plausibly, needs fix
functions. No reason has been given, however, to believe that facts
about needs are naturalistic, and we have good reason to believe
that they are not.

That Dretske fails to take sufficient account of this issue, is
evident from a discussion between Ruth Garrett Millikan and him in
a symposium on his *Explaining Behavior* which appeared in 1990
(Dretske 1990 and Millikan 1990a). Millikan criticizes Dretske,
somewhat as I have here, for failing to give an adequate account of
what it is for something to have a function. She comments that
"Dretske quite studiously avoids committing himself to any
particular notion of (intrinsic) function" (Millikan 1990a, p.809).
Dretske's response to this is to acknowledge openly that Millikan
is right; that he does not attempt to specify in detail what must
be the case in order for something to have a function. His
justification for his silence on this matter is that he considers
this notion to be just one among many —he cites causality,
selection, probability, and learning, among others— which it is
beyond the purview of his work to analyse. Dretske goes on to say,
"all I need is some process in which an indicator acquires a special
status (call it what you will), a status in which there is, among

40

the many things it indicates, some one thing it is now *supposed* to indicate." (Dretske 1990, p.824. Dretske's emphasis). Dretske's locution here is telling. The primary sense in which we understand something's being supposed to do such-and-such, surely, is in its being supposed *by someone* to do such-and-such. That an indicator is supposed by some intentional agent to indicate something, however, is hardly the sort of naturalistic property of the indicator which Dretske needs for his account. The whole point of his account of misrepresentation, indeed, was to explain how it is that states of physical systems can misrepresent the world, independently of any intentional agent's suppositions. Now, Dretske would undoubtedly protest that by "supposed to", he did not mean "supposed by an intentional agent to". The present point, however, is that in this case, what Dretske owes us, but does not provide, is precisely an account of what "supposed to" does mean, in this case.

A second, particularly perspicuous illustration of the strategy which I am criticizing is afforded by Ruth Garrett Millikan's use of "proper functions" in trying to naturalise meaning and rule-following[22]. In a paper responding to Saul Kripke's rule-following scepticism (Millikan 1990b), Millikan proposes to concede Kripke's point, cited in section I above, that rule-following is normative, and yet still to show that it is naturalistic. She says,

> The problem is to account for the *normative* element that is
> involved when one means to follow a rule, to account for there
> being a *standard* from which the facts, or one's dispositions,
> can diverge. (p.329. Millikan's emphases).

She continues,

---

[22]Although some might argue whether the issue is the same where rule-following is concerned and where linguistic meaning is concerned, it is clear from the text that Millikan, like Kripke, takes it that they are.

> In the case of meaning, the normative element seems to be the same as the purposive element: to mean to follow a certain rule is to have as a purpose to follow it. Whether my actual dispositions are "right" or "wrong" depends on whether they accord with what I have purposed. The possible divergence of fact from a standard is, in this case, simply the failure to achieve a purpose. (p.329)

Her strategy, then, is to reduce following a rule or using a word meaningfully to having certain purposes, where these purposes ultimately reduce to 'biological purposes', or what she calls "proper functions" (p.330). As will become evident, there is an equivocation in Millikan's account between a notion of proper function which is genuinely naturalistic, but impotent for her reductivist aims, and a notion which is suitable for her aims, but is not naturalistic.

Millikan originally set out her account of the notion of "proper function" in her *Language, Thought and Other Biological Categories* (Millikan 1984. See Ch.1, esp. p.28). As the exact definition is quite involved, however, and as it is given in terms which are themselves technical terms introduced by Millikan, we will settle for a paraphrase offered in a later paper:

> Putting things very roughly, for an item *A* to have a function *F* as a "proper function", it is necessary (and close to sufficient) that one of these two conditions should hold. (1) *A* originated as a "reproduction" (to give one example, as a copy, or a copy of a copy) of some prior item or items that, *due* in part to possession of the properties reproduced, have actually performed *F* in the past, and *A* exists because (causally historically because) of this or these performances. (2) *A* originated as the product of some prior device that, given its circumstances, had performance of *F* as a proper function and that, under those circumstances, normally caused *F* to be performed by *means* of producing an item like *A*. (Millikan 1989, p.288).

One point worth noting concerning this definition, is that it defines the proper function of *A* in terms of the *function* of *F*, and hence by present standards is not an uncontroversially naturalistic

42

defintion. Since, however, nothing appears to be lost if "the performance of action $F$" is substituted for "a function $F$" in the first sentence of this passage, giving

"... for an item $A$ to have *the performance of action F* as a "proper function", it is necessary (and close to sufficient) that ...",

and since performing an action is a naturalistic matter, I will take Millikan to have meant no more than this. This having been said, what is significant about proper functions is that an item $A$'s having one is solely a matter of $A$'s causal origins, and of the actions performed by $A$'s ancestors[23].

The difficulty with this definition of "proper function" as far as Millikan's naturalistic project is concerned, is that it will be of no help in naturalizing the normative aspect of meaning. If $A$'s having a proper function is solely a matter of its having a certain type of etiology, then $A$'s having a proper function entails no normative consequences —that something *is* descended in certain ways from ancestors of a certain kind, does not entail that the thing *ought* to do anything at all. To see this more clearly, we can consider, ironically, a criticism which Millikan herself directs against a rival account of functions, the disposition or capacity-based account.

Millikan defends her view against accounts which reduce having a purpose or function to having certain current dispositions or capacities, by arguing that having such dispositions or capacities is compatible with not having any purpose or function at all. Millikan takes it upon herself to criticize the dispostion/capacity

---

[23]These comments concerning the definition of "proper function" apply also to Millikan's original definition in Millikan 1984.

account of function because it is a rival to her etiological account, and moreover one which does not have a certain counter-intuitive implication which the etiological account does have. The counter-intuitive implication in question is that it is possible, on the etiological account, for two objects —two (apparent) hearts or carburetors or can-openers— to be physically identical, and yet for one to have a purpose or function and the other not. Thus, suppose that, by an extraordinary freak of nature, a perfect duplicate of me spontaneously came into being, *ex nihilo*, as it were (Cf. Millikan 1989, p.292). According to Millikan's etiological account, the 'heart' in this duplicate would not have as its function to pump my duplicate's blood —it would not have any function or purpose at all, since it would not have the appropriate causal history. On the disposition/capacity account, by contrast, my duplicate would be identical to me in respect of functional properties. To the extent that it is intuitive that my physical duplicate should be my teleological duplicate, then, the dispositional/capacity view is *prima facie* more plausible than Millikan's etiological view.

Millikan's response to this is to bite the bullet, and maintain that my double's heart has no function. She claims that we should not be fooled by the superficial similarity between my double and me; by the fact that its organs have all of the signs of having purposes. The reason why we might be fooled, she suggests, is that nothing close to such a duplicate is ever encountered in reality. Nothing which we might encounter in the normal course of events which has the precise physical structure of, say, a heart ensconced in a body, fails to have the requisite etiology for it to have as

its function, to pump blood. Any dispositions or capacities which a physical system might have, then, are "marks of purposiveness", but that's all they are. Millikan comments,

> [There do not] *in fact* exist complicated goal-directed items, or items displaying complicated negative feedback mechanisms, or items that do anything like "registering" situations, or items with interesting Cummins functions, that do not *in fact* have corresponding proper functions. Having the *right sorts* of current properties and dispositions is in point of *fact*, in *our* world, an infallible index of having proper functions. If you like, it is criterial —as criterial, say, as the red of the litmus paper is of acidity. But it is not turning litmus paper red that *constitutes* acidity, nor is it having the right sort of current properties and dispositions that *constitutes* a thing's having a purpose. To the degree that each of these contemporary descriptions in terms of current properties or dispositions is successful, each describes only a *mark* of purposiveness, not the underlying structure. (Millikan 1989, p.293. Millikan's emphases).[24]

This certainly seems correct. Although it may often be the case that objects which have a certain function also have certain complex dispositional properties, some object's having these dispositional properties is not what constitutes its having the function that it does. It is entirely coherent to suppose that an object has these dispositional properties, but has no function or purpose —in the sense of having an aim or goal— at all.

The rub here, is that there is no reason to think that the situation is any different where Millikan's etiological properties are concerned. Thus, suppose that some object *x* (e.g., a heart or thyroid gland) with property *P* (e.g., a certain mechanical or biological configuration), came to exist in virtue of there having been in the past some object or objects with property *P* which tended to behave in a certain way *b* (e.g., to circulate blood or to secrete

---

[24]In this passage, Millikan appears to be *assuming* that "having a proper function" and "having a purpose" are synonymous. Since this is just what is in question, I take this to be an oversight on her part.

certain hormones), where it was this behaviour of x's antecedents which, in part, caused x to come to be. To a first approximation, this, for Millikan, is a sufficient condition for x's having it as its function to behave in way b. But why should this be so? It is entirely coherent that x should have the etiological properties that it does, and yet have no function at all, if it is being assumed that having a function is a matter of having a role in the fulfillment of some goal. There simply is not any *prima facie* conceptual relation between having a certain causal history, and having a certain goal, which justifies Millikan's assertion of their equivalence. There is nothing in the etiological account, indeed, which impugns the tenability of scepticism about teleological properties in general. This point is reinforced by the very evolutionary theory which inspired the etiological account in the first place. Darwin's great achievement, presumably, was to show that the existence of life on earth can be explained without appealing to any putative teleological facts or principles —that the physical laws and facts acknowledged by the basic physical sciences are sufficient to explain how we got here. The theory of evolution, then, is entirely consistent with scepticism about teleological properties, and hence with the view that the possession of etiological properties is not a sufficient condition for the possession of teleological properties. Insofar as this is so, Millikan's identification of proper functions and purposes cannot be correct. It is of course an option for Millikan to deny that facts about proper functions or 'purposes', as she understands them, are really *teleological*, and entail any normative consequences. If she follows this route, however, then proper functions will be of

46

no use in accounting for the normative aspect of meaning, which she has acceded to.

The last proposal to consider is due to Jerry Fodor. Fodor's paper 'Psychosemantics' (Fodor 1984a)[25] sets out an unabashedly teleological theory of meaning, which is at the same time clearly supposed to be a naturalistic theory. It is worth noting that Fodor published this paper, at the request of an anthologist, only after he had already renounced the theory set forth in it[26]. The reason for going to the bother of criticizing it is that, as the anthologist notes[27], the theory has a measure of currency among some philosophers, despite Fodor's reservations.

'Psychosemantics' takes as its starting point that a certain theory of mind, which Fodor calls 'The Representational Theory of Mind', or RTM, is true. This theory is broadly characterisable in terms of two hypotheses. The first is that token propositional attitudes, such as token believings that *p*, are relations between individuals and token mental representations, where these representations are sentences constructed out of an internal system of symbols —a 'language of thought'. The second hypothesis is that mental processes, such as the drawing of inferences, are just computational operations on sentences of this language of thought, where this is possible in virtue of the syntactic properties of the symbols. Fodor does not purport in 'Psychosemantics' to defend this theory; his project is, rather, to give the framework for a theory

---

[25]This paper should not be confused with Fodor's book of the same name (Fodor 1987).

[26]See the editor's note, Fodor 1984a, p.312.

[27]Ibid.

of meaning for the language of thought on the assumption that there is one.

Given that the mind/brains of cognitive agents are populated with strings of symbols, as RTM assumes, the challenge is to give an account of how these symbol strings come to have the truth conditions which they are supposed to. That is, the question is, what makes these symbol strings *representations*? As Fodor observes (Fodor 1984a, p.317), there are any number of mappings from mental symbols onto states of affairs in agents' environments. The problem is to determine what is special to the mapping which takes symbols onto their truth conditions, which distinguishes it from any other, arbitrary mapping, and makes the symbol strings true of what they are respectively mapped onto:

> ...there is something *special* about the sort of symbol-world correspondence that truth definitions specify; something that shmuth [i.e. any other, arbitrarily chosen] correspondence doesn't have; something in virtue of which facts about their truth conditions illuminate the functioning of symbols in a way that facts about their shmuth conditions do not. Well, yes; but what could this mystery ingredient be? (Fodor 1984a, p.317).

The only constraint which Fodor explicitly makes on any theory specifying this relation, is that it characterize the relation 'Mental symbol M has truth condition S' in non-semantic and non-intentional terms —the point of the theory is not merely to characterize the truth relation in some other semantic or intentional terms. Given Fodor's aim of demystifying the nature of the relation in question, though, an account which explicates it in terms of an equally opaque or mysterious notion of teleology would evidently not be adequate either. What is evidently being sought is a naturalistic account of the relation.

48

According to RTM, some cognitive agent O's having a belief that
S consists in O's standing in some relation R* to some mental
representation M whose truth condition is S. For heuristic purposes,
Fodor suggests that we imagine that for O to bear R* to M is just
for O to have a token of the mental representation M in a certain
memory slot in its brain, a slot which Fodor calls the "yes-box".
Thus, if M has truth condition S and O has a token of M in its yes-
box, then O believes that S. Now, to get an account of what makes
S the truth condition of M, Fodor makes an assumption which he
characterizes as teleological. This assumption is that, among the
various cognitive mechanisms which either add or remove token
representations to or from the yes-box, there is a subset whose
*function* is to add or remove them. Fodor's theory is that it is the
normal functioning of these mechanisms which fix the truth
conditions of mental representations. More exactly,

> The truth condition of O's mental representation M is S iff
> (when O's cognitive mechanisms are functioning normally, a
> token of M is put into O's yes-box iff S obtains[28]).

The key issue here is what it is for a cognitive mechanism to
'function normally'. Fodor suggests that "for heuristic and
expository purposes", we can assume that a cognitive mechanism's
normal function is to such-and-such, if it was by doing such-and-
such that the cognitive mechanism gave to O's ancestors a
selectional advantage (Fodor 1984a, pp.323, 4). The idea, then, is
that what makes S the truth condition of M is that tokens of M being

---

[28]Fodor puts this point in terms of "entry conditions", where
"the entry condition for ... M is that state of affairs such that,
when functioning normally, (the cognitive system brings it about
that M is in the yes box) iff (the entry condition is satisfied)"
(p.324).

put into the yes-box when S obtained conduced to the survival and proliferation of O's ancestors[29].

An example which illustrates the idea is the ubiquitous case of the frog's visual system. According to neurobiologists, when a black spot subtending an angle between 1° and 3° appears in a frog's visual field, certain groups of neurons in the frog's brain are caused to alter their pattern of electrical activity (Lettvin et al., 1959). According to RTM, when this occurs, what happens is that these neurons cause a token mental representation —$M_F$, call it— to be put into the frog's yes-box. The theory above claims that the truth condition of $M_F$ is that there's a fly about iff [when these groups of neurons in the frog's brain are functioning normally —i.e., functioning in that way which, in the frog's ancestors, gave them a selectional advantage— $M_F$ is put into the yes-box iff there's a fly about].

The foregoing is a very brief exposition of Fodor's theory, one which leaves out a number of important qualifications. The precise details of the theory, however, are not at issue here. What is at issue is whether the notion of function which figures in it is really naturalistic, and if it is, whether it is sufficient to reduce meaning facts in the manner Fodor supposes. Given that he "dislike[s] vulgarized Darwinism intensely" (Fodor 1984a, p.323), and hence appeals to evolution only as a heuristic, one might expect Fodor to say more about what in general makes it the case that an organism has the teleological properties which would confer upon its

---

[29]Fodor's insistence that the appeal to evolution to fix functions is merely heuristic, is due in part to his not wanting the truth of his account to rest on the truth of evolutionary biology, as well as to his distaste for what he calls "vulgarized Darwinism".

cognitive systems normal functions. On this point, however, Fodor says merely that he assumes that "the truth (/falsity) of teleological ascriptions has something to do with the truth (/falsity) of corresponding counterfactuals", but admits that he doesn't know just which counterfactuals these are (Fodor 1984a, p.323). His unenthusiastically proffered conception of the evolutionary account of teleology, then, is that the relevant counterfactuals are "facts about natural selection; facts of the type *selection wouldn't have applied in the way it did but that .
. ."* (p.323, Fodor's emphasis and ellipsis). The normal functions of cognitive systems, likewise, are specified in terms of counterfactuals:

> ... for the cognitive system of an organism to be functioning normally is for it to be true of that system that it *would* do such-and-such if it *were* the case that so-and-so ... (Fodor 1984a, p.323).

The general form of the proposal, then, is,

> [Cognitive system S is functioning normally iff S would do *x* if it were the case that *y*] iff [selection wouldn't have applied in the way it did (i.e., giving us organisms with cognitive system S) but that S did *x* when it was the case that *y*]

Now, although Fodor calls the facts about normal functions in question here "teleological", there is nothing in this program for explicating what normal functions and teleological facts generally are, to suggest that possessors of normal functions have aims, and that they are liable to normative evaluation. *Prima facie*, then, it appears that Fodor's strategy is not to appeal to a genuinely teleological notion of function to try to account for the normative nature of meaning, but rather to try to give an account of meaning in entirely non-normative terms. That is, it looks from this passage as though Fodor is not making the problematic assumption of Dretske

and Millikan, that there is a notion of function which is simultaneously naturalistic and normative. In fact, however, it is clear from Fodor's own, later criticism of this account of meaning, that he *does* intend the notion of normal function here to be teleological in the strong sense. As he characterizes the account in his book *Psychosemantics*, the theory just summarized holds that for a cognitive system to be functioning normally is for it to be functioning as it is "supposed to" —i.e., as it would in "optimal" circumstances. He goes so far as to say,

> The teleology story perhaps strikes one as plausible in that it understands one normative notion—truth—in terms of another normative notion—optimality. (Fodor 1987, p.106)[30]

The notion of teleology just cited, then, is evidently *not* a naturalistically innocent notion according to which teleological facts have no normative implications.

The form of the objection to the theory thus characterized should be familiar by now. Fodor's basic claim is that, given a suitable evolutionary history, (cognitive system S is functioning normally) iff (S would do *x* if it were the case that *y*), for a certain *x* and *y*. Now, if S were not an evolved system, then, given that *x* and *y* are merely the initiating of certain brain processes and the obtaining of certain physical facts about the world, respectively, neither S's doing *x* when it is the case that *y* nor S's doing *x* when it is the case that not-*y*, would have normative implications of any kind. But, as has been seen, neither having an evolutionary history, nor having a causal role in the ongoing course of evolution, have

---

[30]For further evidence that the notion of function here is meant to be teleological in the strong sense, see his contemporaneous paper 'Semantics, Wisconsin Style' (Fodor 1984b, esp. p.247)

any normative implications either, by themselves. Presumably, then, 'cognitive system S is functioning normally' is not in any way normative either. But 'functioning normally' is meant to be equivalent to 'functioning as it is supposed to' —i.e., 'optimally'— which *is* taken to have normative implications. Evidently, the normativity in question is being 'smuggled in', rather than deriving from the naturalistic facts at issue. The conclusion, then, is that the source of the attractiveness of teleological meaning naturalism, that it reduces the apparent normativity of meaning to the putative naturalistic normativity of 'normal functions', is specious —there was never any normativity in nature to begin with.

Part of the reason for belabouring this point by going after this long-since rejected theory of Fodor's, is that although this criticism, if it is fair, is fundamental to what is wrong with teleological meaning naturalism, it is not the criticism which Fodor makes. Indeed, he even says that he is "not sure that this teleology/optimality account is false..." (Fodor 1987, p.105), whereas if the criticisms above are valid, it must be false. Fodor's criticism is, rather, that it appears that "there is no guarantee that the kind of optimality that teleology reconstructs has much to do with the kind of optimality that the explication of 'truth' requires" (p.106). The view that meaning naturalism is compatible with the thesis that meaning facts are normative, then, is left unscathed by Fodor's criticism.

As already mentioned, an alternative approach, which Fodor might have taken, is to take the so-called teleological facts mentioned in the account *not* to be teleological in the sense described in Section I above —i.e., to take them not to presuppose that the

organisms in question have any end or goal, relative to which behaviour can be evaluated as right or wrong. On this construal, the matter of whether properties such as having a goal or purpose are naturalistic is not really germane to the present problem, after all, inasmuch as the functions appealed to need not be teleological in the sense of being subordinate to a goal or purpose. All that's really needed, on this view, is that those counterfactuals in terms of which one would analyse the given teleological properties, are true of the relevant cognitive mechanisms. Thus, what's really at issue where meaning naturalism is concerned, is whether meaning properties can be analysed in terms of these counterfactuals.

It was remarked at the outset of this discussion of Fodor, that he renounced the teleological theory set forth in 'Psychosemantics' even before he permitted the paper to be published. The view in favour of which he rejected the teleological theory is essentially the alternative just suggested —that the meaning relation can be reduced directly to counterfactual relations, without any appeal to teleology (See Fodor 1990b). It is worth concluding this chapter by considering how this alternative approach fares, in light of the conception set out in 'Psychosemantics' of what an account of meaning is supposed to accomplish. As it turns out, Fodor finds that teleology can't be eschewed as easily as one might think.

In 'Psychosemantics', Fodor emphasizes that, in trying to give a naturalistic account of meaning, he is not merely trying to give, in naturalistic terms, necessary and sufficient conditions for its being the case that S is the truth condition of M. Rather, his aim is to try somehow to *explain*, in naturalistic terms, how it is that

54

physical objects or states of physical objects can have truth

conditions *at all*. As he says,

> ... for each mental representation M, S is the truth condition
> of M iff S is the entry condition for M. Or rather, something
> stronger: what *makes* S the truth condition for M —the precise
> fact in virtue of which S counts as the truth condition for M—
> is that S satisfy the entry condition[31] for M. (Fodor 1984a,
> p.326. Fodor's emphasis).

The suggestion here is that it is not merely the case that, as a

matter of fact, S is the entry condition of M iff S is the truth

condition of M, but rather that S's being the entry condition of M

somehow explains why S is the truth condition of M. That is, S's

being the entry condition of M is Fodor's "mystery ingredient"[32]

which accounts for why, among all of the various states of affairs

in the world to which M is related, S is the truth condition of M.

The point is put again toward the end of 'Psychosemantics':

> After all, the project is not to *analyze* 'S is the truth
> condition of M'. If that had been the project, we would have
> had considerable inductive grounds for giving up at the onset.
> Analyses are what close out of town; surely it is the lesson
> of modern philosophy that interesting ideas don't get
> analyzed. No, the point was rather to show *how it is possible*
> that mental symbols should have truth conditions even though,
> since there are no agents who employ them, their semanticity
> cannot trace back to the intentionality of anybody's
> propositional attitudes. (Fodor 1984a, p.333. Fodor's
> emphases.)

In short, what Fodor is after is some set of naturalistic conditions

which in some sense 'add up' to some state of affairs' being the

truth condition of some symbol.

Now, whatever their shortcomings, teleological theories have at

least some *prima facie* appeal as a solution to this problem. Roughly

speaking, according to these theories, the "mystery ingredient"

---

[31]Cf. footnote 28.

[32]Cf. the passage cited above, from Fodor 1984a, p.317.

which makes a mental symbol M a candidate for being true is that it has the purpose of indicating, to a certain goal-seeking cognitive system, that the world is in a certain state, which is perhaps plausible. Moreover, as has been seen, teleological theories appear, at least initially, to be able to explain convincingly how it is that a mental symbol can *mis*reprepresent the world: if system X has as its (only) purpose to bring it about that Y, then it follows that any behaviour of X's which decreases the likelihood of Y's coming to obtain is a *mistake* or *error*. Misrepresentations, perhaps, are just errors of this kind.

The teleological construal of Fodor's 'Psychosemantics' account of meaning has at least this much intuitive plausibility, then, as a solution to the problem of naturalizing meaning as it has just been construed. But what of the theory which is being proposed as an alternative, which would obviate the charge of being non-naturalistic simply by dropping appeal to teleology altogether? On this alternative account, the "mystery ingredient" which makes a mental symbol a candidate for being a representation of the world is merely its standing in a certain counterfactual relation to the world. But does this account show "how it is possible that mental symbols should have truth conditions?" It is not obvious to me that a physical object or state, merely by having its token occurrences in a certain way counterfactually related to some other physical object's occurrences, thereby possesses *truth conditions*. More to the point, Fodor himself appears to have doubts on this point, as is evident from his evaluative comments on his alternative theory. At the close of 'What is a Theory of Content, II' (Fodor 1990b), Fodor asks, in effect, whether, after all, some mental symbol's

standing in the appropriate counterfactual relations to the world, adds up to its having content; whether the account is sufficient to "solve Brentano's problem"[33]. One objection to a positive answer to this question for which Fodor admits some sympathy, is that "we want more" than merely that some mental symbol stands in a certain counterfactual relation to some state of affairs. As he says,

> "...we want it to turn out that some ways of using symbols are *wrong*. Where, in the picture of representation that we've been constructing [i.e., the non-teleological, counterfactual picture], does the idea get a foothold that there are *mis*representations; and that they are things to be avoided?" (Fodor 1990b, p.128)

Since Fodor has admitted that one of the virtues of the teleological theory is that "it's sensitive to the intuition that errors are cases where *something has gone wrong*" (Fodor 1990a, p.64), he ought to have something to say about the failure of the counterfactual theory to address the intuition. Well, what does he say? As it happens, what Fodor suggests in the end, admittedly with some trepidation, is that what is needed to meet the intuitions is some marriage of his counterfactual account with suitable teleological considerations. In other words, he thinks that (perhaps) we will need teleology after all, to account for misrepresentation.

The moral of this lengthy digression in response to the suggestion that Fodor doesn't really need any teleology in his account, is that the purposiveness of the cognitive mechanisms appealed to in the initial theory plays a substantive role in the

---

[33]Specifically, what Fodor asks is, "...does information and robustness equal content? Are information and robustness all you need for intentionality?" (Fodor 1990b, p.128). What Fodor argues in this paper, is that any sign M standing in the given counterfactual relation to some state of affairs S conveys information about S, and does so "robustly", i.e., (roughly) in such a way that tokens of M can be variously caused (Cf. Fodor 1990b, p.91).

theory, and cannot be dispensed with unless compensatory measures are taken. And, since Fodor gives no account of how the teleological properties in question are any more naturalistic than the meaning properties they are supposed to explicate, I conclude that he fails to naturalize meaning, or even to give any hope that the project is practicable.

# CHAPTER 2: TWO CONCEPTIONS OF THE PROBLEM ABOUT RULE-FOLLOWING

In chapter VI, §38 of *Remarks on the Foundations of Mathematics* (Wittgenstein 1978), Wittgenstein comments

> "How can one follow a rule?" That is what I should like to ask.

Now Wittgenstein does not ask, and nor have philosophers generally asked, "How can one climb a flight of stairs?" or "How can one eat an apple?", so one might wonder just why the activity of following a rule, in contrast to these other activities, should be thought problematic. One possible reason is that it appears to be the case that which rule a person is following on an occasion is underdetermined by any fact or facts about her and her surroundings. This is the problematic aspect of rule-following which Saul Kripke discusses in his *Wittgenstein on Rules an Private Language* (Kripke 1982). A second possible reason is that it appears to be the case that in order to be able to follow a rule, one must be able to justify continuing it in one way as opposed to another, but it turns out that one cannot. This problem is suggested by some of Wittgenstein's remarks, and it is entertained, although ultimately not pursued, by Kripke as well. The difference between these two conceptions of what makes following a rule problematic, I will argue, is substantial. What emerges when the difference is elucidated is that whereas the first problem about rule-following is a problem only for defenders of the view that facts about rule-following reduce to some other kind of facts, the second problem is a problem for anyone who maintains that she follows rules.

This chapter is in three parts. In the first part, the two arguments just adumbrated for the conclusion that one cannot follow

a rule, will be set out. As it may be tempting to think that the difference between them is merely cosmetic, this will be done in some detail. In the second part, the proposal that to be a follower of a rule is just to have a certain set of dispositions will be considered, and the refutations of this proposal suggested by the two arguments, contrasted. The difference between these refutations will serve to illustrate the extent of the difference between the two conceptions of what is problematic about following a rule. In the final part, a brief account of the solution to the problem posed by the latter argument will be sketched.

I

We begin by discussing the argument for the conclusion that the facts underdetermine which rule a person is following. Although the argument as it will be formulated here is drawn from the argument which Saul Kripke gives in his *Wittgenstein on Rules and Private Language* (Kripke 1982, p. 7ff), it is different in small ways. The significance of the minor divergences will be considered below. The argument can be broken into two steps. First, consider some arbitrary rule-following activity which our friend Jones has apparently pursued on many occasions in the past, such as counting by twos —i.e., following the rule described by the formula $a_{n+1} = a_n + 2$ $(a_0 = 2)$. As with any rule, of course, Jones will necessarily have followed it only finitely far to date. For the sake of argument, then, suppose that he has not hitherto contemplated any instance of the rule greater than $a_n = 20004$. Next, consider a peculiar activity which, for convenience, we can assign the name 'counting* by twos', which is described by the following formula:

$a_{n+1} = a_n + 2$ if $a_n$ is less than 20004, but $a_{n+1} = a_n + 4$ otherwise. Someone who was counting* by twos, then, would write "...19998, 20000, 20002, 20004, *20008, 20012,* ..." (Cf. Wittgenstein 1978, p.36). We are led to ask the following question: Given that in the past, whenever Jones engaged in the activity which we would call 'counting by twos', his behaviour was in fact consistent with his having been counting by twos and with his having been *counting** by twos, what further fact is there about him or his environment which makes it the case that he was doing the former rather than the latter (or vice-versa), and as a result of which he should follow 20004 with 20006 rather than 20008 (or vice-versa), to accord with his past intentions?[1]

The second step of the argument is an examination of what are presumed to be all further facts about Jones which, coupled with his actual behaviour, might determine which of the two rules he has been following. In each case, an argument is given to the effect that the given fact does not suffice to determine which rule Jones has been following. Among the candidates which Kripke weighs and rejects are facts about Jones' dispositions, facts about any mental experiences or images which Jones may have had, and considerations of simplicity. The conclusion of the argument is that there is no fact as to which rule Jones has been following. If this is so, then, since peculiar rule-following hypotheses such as the hypothesis that Jones has been 'counting*', can always be contrived, and since the

---

[1]Kripke, of course, constructs this argument in terms of his own past behaviour and the rule for giving $x$ plus $y$, on the one hand, and the rule for giving $x$ 'quus' $y$ on the other, where $z = x$ quus $y$ iff ($x$, $y < 57$ and $z = x + y$, or $x$, $y \geq 57$ and $z = 5$) (Kripke 1982, p.9). The argument has been reformulated here to illuminate the comparison with the second argument, developed below.

argument can be generalized to anyone, it follows that there is never a fact as to which rule a person is following on an occasion.

The argument is presented here in terms of rule-following, but it can be extended to apply to meaning something by an expression. First, suppose that it is given that, when Jones says, on an occasion, e.g., "That's green", there is a fact as to what it is that Jones is using the predicate 'green' to describe. This supposition has no obvious analogue in the rule-following case, but inasmuch as denying it can only help the sceptic, there can be no objection to making it here. Next, consider the set of all things which Jones has heretofore used the predicate 'green' to describe, and consider the property of being *grue*[2] —i.e., the property of being green and observed before some future date, or blue and not observed before the date[3]. In a manner analogous to Jones' past counting behaviour, his uses to date of the predicate 'green' to describe things, considered alone, are consistent with his having meant *green* by it and also with his having meant *grue* by it, since all objects encountered to date which have had one property will have had the other. And, as was the case with rule-following, a case can be made that no further fact about Jones determines which of the two properties he meant.

Since it is not primarily my purpose to evaluate this sceptical argument but rather, to contrast it with a second, related argument, I will not go through Kripke's arguments against each of the

---

[2]See Goodman 1983, p.74.

[3]Cf. Kripke's example, that my uses of the word 'table' are consistent with my having meant by it the property of being a table or the property of being a *tabair*, where something is a tabair iff it is a table and not at the base of the Eiffel Tower or is a chair and is at the base of the Eiffel Tower. (Kripke 1982, p.19)

suggestions as to what fixes which rule Jones has been following or what Jones means by a given predicate[4]. Concerning Kripke's general strategy, however, there is a point worth noting here. Both Colin McGinn and Paul Boghossian rightly criticize Kripke for assuming more or less uncritically that facts about rule-following and meaning, if such there be, must reduce to some other kind of facts and, moreover, that these facts must be non-semantic, non-intentional facts (see McGinn 1984, p.150-1, and Boghossian 1989, pp.540). As McGinn says, "[Kripke's] sceptic is in effect demanding an answer to the question 'what does meaning/reference consist in?' which does not just help itself to the notions of meaning and reference" (McGinn 1984, p.151). Admittedly, Kripke considers the possibility that to follow a given rule or mean something by a given word is to be in an irreducible, "*sui generis*" state, and even goes so far as to concede that the suggestion that this is the case "may in a sense be irrefutable" (Kripke 1982, p.51). As both McGinn and Boghossian observe, however, Kripke's cursory dismissal of this suggestion on the grounds that it leaves the nature of meaning "completely mysterious", is not sufficient to show that it is false. And, if there *are* irreducible facts of the form 'Jones is following the rule 'add two'', then, provided some such fact is true of Jones, Kripke does not get his conclusion. For future reference, call the thesis that facts about rule-following and meaning, if there are any, must be reducible to facts of some other kind, the 'Reductivist Thesis'. Whatever its plausibility, it is clearly a substantive thesis which is assumed to hold true in the second step of the argument.

---

[4]The suggestion that it is facts about Jones' dispositions, however, will be taken up in Section II, below.

For future reference, this argument can be encapsulated thus:

(1) Since i)there are always different rules which are consistent with any given (apparent) rule-following behaviour of any person, and ii)assuming that facts about rule-following must reduce to some other kind of facts, then, since iii)there are no other facts about any person which determines which, among these various different rules, she or he has been following, it follows that iv)there is never any fact as to which rule a person is following on an occasion.

There are two aspects of argument (1) which are worth taking note of. The first is that it concerns the existence of *facts* about rule-following or meaning. The challenge posed by the ‘counting*’ and ‘grue’ scenarios is to come up with some account of what following one rule rather than another, or meaning one property rather than another, consists in —what fact or facts about Jones constitute his being a follower of one rule as opposed to another or his meaning one property as opposed to another. In this respect, it is different from an argument concerning our *justification* for believing that Jones is following one rule rather than another, or from an argument concerning Jones’ justification for believing that he is following one rule rather than another[5]. One could, after all, give an account of what facts constitute Jones’ following one rule as opposed to another, and yet fail to give an account of what justifies anyone in believing that Jones is following one rule as opposed to another. The sort of account of meaning briefly outlined by Paul Boghossian at the end of ‘The Rule-Following Considerations’, which he calls ‘robust realism’, illustrates this point. According to this account, "judgements about meaning are factual, irreducible, and judgement-independent" (p.547). That

---

[5]McGinn makes this point in his discussion of the present argument —he calls its conclusion "a constitutative or metaphysical claim", and distinguishes it from an epistemological argument also suggested by Kripke (McGinn 1984, p.149).

someone means such-and-such by a given word, on this view, is not a matter of some non-semantic, non-intentional meaning-constituting facts being true of her, nor is it a matter of her *being judged* to mean such-and-such by some well-placed interpreter(s) —it is, rather, a matter simply of a certain irreducible meaning fact's being true of her. What Boghossian is proposing, in effect, is that the Reductivist Thesis is false, and that meaning facts are *sui generis* —a view which, as remarked just above, Kripke perfunctorily dismisses. Given that the sort of meaning-properties which this view posits do not, by hypothesis, reduce to any properties which our (commonly acknowledged) perceptual faculties are attuned to detect, the basic theory will leave the epistemological questions just noted, unanswered.

It may be remarked here that whereas argument (1) has this noteworthy aspect —viz., that it concerns the existence of facts about rule-following and not matters of justification— the argument which Kripke gives in *Wittgenstein on Rules and Private Language*, does not. Thus, whereas Kripke does maintain that his argument concerns what fact makes it the case that, in the past, he was following such-and-such rule, he also asks what justifies him in going on as he does. Kripke says,

> ...the challenge posed by the sceptic takes two forms. First, he questions whether there is any *fact* that I meant plus, not quus, that will answer his sceptical challenge. Second, he questions whether I have any reason to be so confident that now I should answer '125' rather than '5' [to the question "What is 68 + 57?"]. (Kripke 1982, p.11)

Later in his exposition, Kripke construes the second form of the sceptic's challenge in more clearly epistemological terms, as a demand for *justification* for giving one answer rather than the other (see, e.g., p.15). Kripke, then appears to be concerned with more

than just what fact makes it the case that one rule as opposed to another is being followed.

This point about Kripke's exposition must be granted. The principal aim of this chapter, though, is to show that what Kripke takes to be two forms of one argument, are in fact two substantially different arguments.

The second noteworthy aspect of argument (1) concerns the epistemological perspective of the meaning theorist (turned sceptic). Roughly speaking, the point is that he has a third-person, rather than a first person, perspective on the problem. In formulating the question posed in the first step of the argument, the meaning theorist contrasts the possibilities that Jones has been counting by twos and that Jones has been counting* by twos. And, in the second step of the argument, the meaning theorist evaluates various candidate facts to see whether they determine which, if either, of these possibilities obtains. In order to be able to determine which possibility obtains, the theorist must know what the two possibilities are —he must know, that is, what it is to count and what it is to count*, by twos. If he did not know this, he could not hope to be able to judge whether a given fact determines that Jones has been doing the one as opposed to the other. In this respect, the meaning theorist is different from Jones, his object of study. Jones, after all, may turn out to have been neither counting nor counting*, nor to have been following any other rule. In this case, Jones need not know either what it is to count or to count*. It may turn out, indeed, that Jones has no intentional states at all, and hence does not *know* anything. Clearly, then, there is an asymmetry between the meaning theorist's possible

epistemological states and Jones's. It should be noted that, if the meaning theorist arrives at this conclusion —that Jones has no intentional states— then the initial assumption that *he* knows what it is to count and to count*, must be questioned. For, argument (1) puts no constraints on Jones' constitution, whence its conclusion is not affected if it is assumed that Jones and the meaning theorist are, in all relevant respects, identical. In this case, the meaning theorist has to reject the initial assumption that he knows what it is to count and what it is to count* —he has to "kick away the ladder" (Cf. Kripke 1982, p.21)[6].

This second point about argument (1) is also true of the argument as it appears in Kripke's exposition, although this is perhaps obscured by certain differences in the formulations. Rather than inquiring, concerning some arbitrary person (here, Jones), whether he has been following one rule as opposed to another, Kripke inquires, concerning his own past self, whether he meant plus or quus by 'plus' (Cf. footnote 1). It may look, then, as though Kripke poses the problem from a first-person, rather than a third-person point of view. The point of his asking about his past self from the perspective of his present self, however, is precisely to distinguish the person, the meanings of whose utterances are in question, from the person inquiring about the utterances. As Kripke says,

> I put the problem in this way so as to avoid confusing questions about whether the discussion is taking place 'both inside and outside language' in some illegitimate sense. If we

---

[6]This is the position in which Kripke's sceptic finds himself, in the end. It should perhaps be noted that a more accurate metaphor would be that the sceptic has to deny the existence of the ladder climbed up: he has to deny that, in making his argument, he expressed any propositions.

are querying the meaning of the word 'plus', how can we use it at the same time? (Kripke 1982, p.12)

For Kripke, then, the problems 'What makes it the case that in the past I meant plus by 'plus'?' and 'What makes it the case that Jones means plus by 'plus'?' are not relevantly different. And, as we have seen, the construal of the argument in terms of word-meaning is parallel in all relevant respects to the argument about rule-following.

In respect of these two points, step 2 of argument (1), in which a search is undertaken for some fact which determines which rule Jones has been following, resembles the type of theorizing about knowledge and belief called 'naturalized epistemology'. According to this conception of epistemology, it is merely a branch of psychology which has as its aim to determine what relation a cognitive agent must stand in to his environment, and by what causal route the agent must have passed to be in the relation, in order for the agent to stand in some epistemic relation to some fact[7]. The problem posed to the practitioner of naturalized epistemology, like the problem posed in step 2 of argument (1), is to come up with some fact or facts about some person which make it the case that he is in a particular state —in the case of naturalized epistemology, the state of knowing or being justified in believing that *p*; in the case of the meaning theorist in step 2 of argument (1), being a follower of a given rule. And the practitioner of naturalized epistemology, again like the meaning theorist in step 2 of argument (1), poses her question from a detatched, third person point of view. That is, she

---

[7]See, eg., Quine 1969, pp.82-84, and Alvin Goldman, *Epistemology and Cognition* (Cambridge, Mass.: Harvard, 1986), esp. chapters 4 and 5.

is assumed to know certain important facts which it is possible that her object of study does not. The naturalized epistemologist, after all, is assumed to know the state of Jones' cognitive machinery, as well as whether the proposition *p* which Jones is supposed to believe (say), is true. Otherwise, she can hardly be expected to come up with a reliable assessment of Jones' cognitive state. It is quite possible, however, that it should turn out that Jones himself has no knowledge of, nor justified beliefs about, the state of his cognitive machinery or the relevant state of the world. Thus, like the meaning theorist, who is assumed to know facts about what it is to follow certain rules which his object of study may not, the naturalized epistemologist is assumed to know facts about her object of study's cognitive states and about the state of the world which her object of study may not.

Argument (1), then, represents the first reason for thinking that there is something problematic about following a rule. A second argument is strongly suggested by various passages in Wittgenstein's writings, and is also suggested by certain comments of Kripke's[8,9]. In his *Remarks on the Foundations of Mathematics* (Wittgenstein 1978), Wittgenstein asks,

---

[8]The following argument is in some respects similar to the characterization of the rule-following problem as a dilemma, in McDowell 1984. See section III below for more on this.

[9]In what follows, I do not mean to be defending an interpretation of Wittgenstein. As it happens, I believe that the following argument reconstructs Wittgenstein's thinking better than the argument given above, but I will not argue for this claim.

How can the word "Slab"[10] indicate what I have to do, when after all I can bring any action into accord with any interpretation?

How can I follow a rule, when after all whatever I do can be interpreted as following it?

What must I know, in order to be able to obey the order? Is there some *knowledge*, which makes the rule followable in only *this* way? (p.341 —Wittgenstein's italics)

The problem is put even more directly in the following passage:

*How do I know* that in working out the series +2 I must write
                    "20004, 20006"
and not
                    "20004, 20008"?
—(The question: "How do I know that this colour is 'red'?" is similar.)".
    "But you surely know for example that you must always write the *same* sequence of numbers in the units: 2, 4, 6, 8, 0, 2, 4, etc." —Quite true: the problem must already appear in this sequence, and even in *this* one: 2, 2, 2, 2, etc. —For how do I know that I am to write "2" after the five hundredth "2"? i.e. that 'the same figure' in that place is "2"? And if I know it *in advance*, what use is this knowledge to me later on? I mean: how do I know what to do with this earlier knowledge when the step actually has to be taken? (p.36 —Wittgenstein's italics).

These passages suggest that I am never justified when, in doing what I take to be following a rule, I continue in one way rather than another. To spell this out, imagine that I am seated at my desk, engaged in what I would normally take to be some rule-following activity —counting by twos, let us suppose— writing down successive values of the series more or less automatically. In a moment of philosophical self-examination, however, I ask myself "What justifies me in writing down '20006' and not '20008' after '20004'? —What is my ground for writing one thing rather than another?". In the normal course of events, evidently, I simply write down the successive numbers unthinkingly, but I believe that I am doing more

---

[10]Cf. the language game in which a builder shouts "Slab!" to his helper in *Philosophical Investigations* (Wittgenstein 1958), §2.

than merely writing down whatever strikes my fancy. The question raised is an epistemological one: although I surely seem to be justified in believing that 20006 follows 20004 in the given series, the question is what my justification *is*.

One possible answer to this question is that I am justified in following 20004 with 20006 rather than 20008 (or anything else), because when I count by twos I follow the more basic rule 'count by ones and write down every other number', and this specifies that 20006, and not 20008, follows 20004. As Wittgenstein's familiar comments on rules for following rules make plain, however, this will not suffice[11]. This answer simply raises the questions 'What justifies me in following 20004 by 20005 rather than anything else when counting by *ones*?', and 'What justifies me in following the rule "Write down every other number" as I do?'. In the absence of answers to these questions, I have no justification for succeeding 20004 with 20006 in following the initial rule. This, moreover, indicates the pervasiveness of the problem: if there is a problem about my following such basic rules as the rule for counting by ones, then there is likewise a problem about the whole edifice of my rule-following activities.

Related objections attend the suggestion that I am justified in writing one number rather than another by some sort of intuition. As Wittgenstein notes,

> If intuition is an inner voice —how do I know *how* I am to obey it? And how do I know that it does not mislead me? For if it can guide me right, it can also guide me wrong.
>     ((Intuition an unnecessary shuffle)). (Wittgenstein 1958, §213)

---

[11]Cf. Wittgenstein 1958, §§86, 152, 211, 239.

72

It is not clear how an intuition could *justify* any given continuation of the rule, unless there is reason to believe that the intuition is a reliable guide. How do I know that my intuitions are not systematically wrong?

The outcome of these intial reflections is that I have no justification for continuing any given rule, at any juncture, in any way. This is as much as to say that in continuing any given rule, I ultimately must act without the support of reasons —I must make a 'leap in the dark'. To borrow a comment of Wittgenstein's, "When I obey a rule, I do not choose. I obey the rule *blindly*"[12]. That this is the case, is the first premise of this second argument:

> Premise 1: In doing what I take to be following a rule, I have no justification for continuing, at any juncture, in any particular way.

> Consider next, the following platitude about rule-following:

> (P) If I am following a rule, then, assuming I do not want to make a mistake, I am constrained at any given juncture to continue in a particular way —I am not free to go on as I please.

If I am counting by twos and I do not want to make a mistake, then after I've written, say, "546", I am constrained to write "548". More generally, if, in writing down symbols on a page, I were free at any juncture to write whatever I pleased, then I would not be following a rule —what I would be doing would best be described as doodling. This platitude, however, very plausibly entails

> (P') If *I am justified in believing* that I am following a rule, then *I am justified in believing* that I am constrained at any given juncture to continue in a particular way.

---

[12]Wittgenstein 1953, §219.

How, after all, could I be justified in believing that I am following a rule, without my thereby being justified in believing that I am constrained in how I can go on?

Observe next, however, that if, in doing what I take to be following a rule, I have no justification for continuing, at any juncture, in any way —if, that is, Premise 1 is true— then it appears to be the case that I am *not* justified in believing that I am constrained as to how to continue. For it appears that nothing could count as justifying my belief that I am constrained to go on in a particular way in pursuing some activity of mine, other than my having the sort of justification for carrying on in one way as opposed to another at any given juncture, which Premise 1 denies I have. To put this another way: prior to having reflected on any of the foregoing, I would presumably have said that my justification for believing that I'm following a rule on a given occasion is that I have some justification for going on in particular cases as I do. In the absence of my having any justification in particular cases, however, it appears that I cannot justify my belief that I'm constrained to carry on in a particular way. For future reference, call this claim (C):

> (C) If, in following a rule, I have no justification for continuing at any given juncture in any way, then it is not the case that I am justified in believing that I am constrained at any given juncture to continue in a particular way.

But from Premise 1, (C), and P′, it follows that I cannot justify my belief that I'm following a rule rather than merely producing arbitrary noises or scribbles. If this is the case, then my belief that I'm following a rule on any given occasion is really just a ground-level, unjustifiable conviction.

The second Premise of this argument, then, is that if I have no justification for carrying on in one way rather than another in doing what I take to be following a rule —that is, if I carry on 'blindly'— then I have no justification for my belief that I am following a rule and not just speaking or writing arbitrarily[13]. Again, if this is correct, and if the first premise is also correct, then I am never justified in believing that I'm following a rule.

One might object to this that what justifies one's belief that one is following a rule is some sort of 'feeling'. This objection is inspired by the thought that, in following a rule, one has something like a feeling of concentration or of compulsion which is absent when one merely produces arbitrary noises or scribbles. This suggestion, however, misses the point. What I lack is a reason as to *why* those experiences which I take to be experiences of following a rule are evidence that I *am* following a rule. These experiences of concentration or compulsion give me no justification for believing that I am constrained in how to go on by anything other than my transient preferences.

As it stands, this conclusion seems to threaten my apparently reasonable belief that I sometimes follow rules. Before proceeding to consider what the threat amounts to, it is worth considering one prima facie plausible response to it. This response is to accept the conclusion, but to deny that beliefs about rule-following, in order to be reasonable, must be justified. That is, the suggestion is that my belief that I am following a rule is reasonable despite my lacking any justification for it. Now, insofar as the principle that

---

[13]This premiss is just the conjunction of (C) and the contrapositive of (P').

a belief can be reasonable although not justified, is appealed to in responding to scepticism about certain kinds of facts other than facts about rule-following, this possibility looks promising. For example, where my belief that there is an external world is concerned, plausibly, the answer to the sceptic who argues that it is unjustified, is to *agree* that I have no justification, but to maintain that the belief is reasonable nevertheless. If forms of scepticism such as scepticism concerning the existence of an external world can be met by appealing to this principle, then perhaps it applies where rule-following is concerned, as well.

The claim in question, then, is that my belief that, in saying or writing what I have been, I have been following a rule, is reasonable although not justified. Now, there are three conceptions of what the content of the belief might be, each of which entails one or another problem for this claim. First, it might be supposed that any arbitrary sequence of numbers (or letters or whatever), however erratic, is the extension of some rule, and that the content of my belief that I am following a rule is merely that I am following some one such rule. In this case, the difficulty is that it will be a matter of indifference whether my belief is true or not. Regardless of what I say or write, there will always be a rule to which I have conformed with complete accuracy —indeed, for any finite sequence, there will always be infinitely many rules to which it conforms (given that any finite sequence is a part of infinitely many infinite sequences). In such a case, then, the only difference between my acting randomly and my following a rule, would be the absense or presence of an epistemologically inaccessible relation between my actions and the rule. Since my belief that I'm following

a rule is presumably not supposed to be irrelevant in this way, this cannot be what the content of my belief is supposed to be.

A second suggestion is that only certain sequences —'simple' or 'natural' sequences— are the extensions of rules, and that the content of my belief is that I am following a rule thus construed. The problem with this suggestion can be put as a dilemma. If, on the one hand, what counts as a simple or natural rule is determined by me, then my belief that I'm following a rule becomes a trivial matter. If it is up to me to decide what is and what is not simple, then whether I'm following a rule —i.e., whether the sequence I've been producing conforms to a 'simple' rule— is up to me. The problem with this suggestion is that, if it were correct, then "thinking one was obeying a rule would be the same thing as obeying it" (Wittgenstein 1958, §202). On the other hand, if what counts as a simple rule is fixed by something independent of me, then my having merely a brute conviction that I'm following a rule cannot be enough to make the belief that I am, reasonable. For, if the rule is truly independent of me, then the question of whether or not I'm following it will not be resolvable by appeal solely to how things stand with me. The question of whether or not my brute convictions are reliable i: licators of whether I'm following rules, that is, will not be resolvable solely by looking at my brute convictions.

A third and final suggestion is that the content of my belief is that I'm following a particular rule, which I have somehow identified. The difficulty with this suggestion is that it raises the question as to how I can identify the particular rule in the requisite way. For how can I specify —to myself or anyone else— which rule I'm following, other than by writing out its instances?

As Wittgenstein repeatedly emphasizes, it is no help for me to appeal to a formula or picture here, inasmuch as what is at issue is what the formula or picture prescribes[14]. It is precisely the difficulty of giving content to the idea of having a rule 'in mind' independently of giving instances of it, which gives rise to this rule-following problem in the first place.

The upshot of these considerations is that the conclusion reached above, that I'm never justified in believing that I'm following a rule, entails the conclusion that it is never reasonable for me to believe that I'm following a rule. If the basic argument is correct, then, it follows that every belief of mine that I am following a rule is an unreasonable belief, which, if I ought not to believe what is unreasonable, I ought to renounce. This is not, of course, to deny that what I *call* 'following a rule' 'feels' different than, say, composing frivolous free-form nonsense poems. The point is, rather, that my following a rule is supposed to be distinguished by its being something which constrains me in how I can go on, and it appears that I can have no reason to think that those activities which I have hitherto counted as following rules have this property. If this is correct, then the distinction between rule-following and speaking or writing arbitrarily, for me, is idle. If being correct or incorrect is what distinguishes instances of following a rule from instances of following a whim, then I have no grounds for drawing this distinction —I have no justification for taking there to be such a thing as following a rule. If I believe only what it is reasonable to believe, then I will conclude that there is no such thing as following a rule. This argument, then, shares in common

---

[14]See, for example, Wittgenstein 1958, §§86, 152, 198, 201.

with the first argument that it finds that the notion of following

a rule is empty.

The foregoing argument, then, can be briefly encapsulated thus:

> (2) When engaged in activities which I am inclined to call 'rule-following' activities, I have no justification at any given juncture for continuing in one way as opposed to another. If this is the case, however, then I can provide no justification for my belief that I am ever following a rule and not just acting or writing arbitrarily. Since my belief that I'm following a rule is the sort of belief which is reasonable only if it can be justified, it is not reasonable for me to believe that there is a distinction between my following a rule and speaking or writing arbitrarily.

As with argument (1), this argument applies also to meaning

something by a word. To see this, imagine that, instead of sitting

at my desk, I am sitting at a window picking out green automobiles

as they pass by outside, and I ask myself 'What justifies me in

calling that last Toyota 'green' but not this Chrysler?'[15] The

question this time concerns what justifies me in using a word as I

do. I have used the word 'green' in the past on countless occasions,

granted, but how comes it to be that I am now justified in applying

it as I do? It cannot be simply in virtue of my having in mind a

mental colour sample, inasmuch as the question 'What justifies me

in counting this colour and the colour of my mental sample as the

same?' would be no less problematic than the question at hand[16].

The question, then, is, what justifies me in calling this object

'green' but not that, and if nothing does —that is, if my uses of

'green' are blind— then what is the difference between my using the

predicate 'green' meaningfully and my merely chattering arbitrarily?

---

[15]That is, I'm doing what I would, prephilosophically, take to be picking out green automobiles.

[16]Cf. Wittgenstein 1958, §73.

There are two aspects of this argument worth taking note of, parallel to the two aspects noted of argument (1). The first is that it concerns my *justification* for taking myself to be following a rule, rather than what fact makes it the case that I am a follower of some rule. Notice, again, that an account of the latter would not necessarily yield an account of the former —there might be a fact as to whether I'm in some state without my being justified in taking myself to be in the given state. The second point is that the person whose justification for taking himself to be following a rule I am calling into question, is *myself*. That is, the inquirer and the subject of inquiry are the same person[17]. This makes the argument quite different than, say, the argument to the weaker conclusion that I am not justified in taking my friend Jones to be following a rule. Such a conclusion might be true even if the conclusion of argument (2) is false. If, on the other hand, I am not justified in taking myself to be following a rule, then nor can I be justified in believing that anyone else is following a rule. If I cannot distinguish my following a rule from my chattering arbitrarily, then I have no rational basis for maintaining that I ever follow rules, and if this is the case, then nor do I have a rational basis for forming judgements as to whether anyone else ever does. In these two respects, argument (2) differs from argument (1). Argument (1), it will be recalled, concerned whether there is any fact as to whether some person distinct from my present self is following a rule.

---

[17]In terms of Jones, then, the point would be that it is Jones' justification for taking himself to be following a rule, rather than our justification for taking Jones to be following a rule.

In the two respects just noted, argument (2) resembles traditional epistemological arguments such as Hume's argument that he has no justification for believing in bodies whose existence is continued and distinct from his own. His argument is that any belief in the existence of such a body would have to be justified either by sensory evidence or by reason, and neither suffices. The senses are insufficient, because they afford evidence for the existence of a body only as long as the body is being sensed, and because they deliver at best 'impressions' from which the distinctness of bodies must be inferred, rather than direct evidence of this distinctness (Hume 1739, Bk I, Part IV, §II). Supplementing Hume somewhat, we can say that reason is insufficient because it affords no principle which implies the existence of distinct, continuous bodies[18]. Like argument (2), then, Hume's argument (as supplemented) trades on what he, given his epistemically limited perspective, is *justified* in believing. Hume does not purport to show that there can be no bodies whose existence is continued and distinct from his own; rather, he argues that he cannot justify believing that there are any. And, again as with argument (2), Hume deduces this from what *he himself* is justified in believing —an inquiry into what his friend Jones is justified in believing, evidently, would presuppose the existence of Jones, which the initial conclusion precludes.

On the model of this argument of Hume's, we might say that the conclusion of argument (2) is that my belief that I am following a

---

[18]Hume in fact argues, not that reason *cannot* justify the belief in the existence of continued and distinct objects, but that reason is at least usually not the source of the belief. He concludes this on the grounds that "the greastest part of mankind" have no knowledge of any arguments that might be given for the existence of continued and distinct objects, but believe in them nevertheless. (Ibid.)

rule as I write "2, 4, 6, 8, ..." is neither the product of my senses nor of reason, but of my imagination (Cf. Hume 1739, p.244). Understood in this way, it is apparently quite different than the conclusion of argument (1). Indeed, arguments (1) and (2) are as different as are the arguments which arise in naturalized and Humean epistemology.


## II

Arguments (1) and (2), then, appear to be different arguments. It may remain unclear, however, just what their differences come to. In order to illuminate this, it will be useful to compare what each of the arguments suggests is inadequate with some particular theory of rule-following or meaning. The obvious choice for such a theory is the dispositional theory, which Kripke criticizes in the course of setting out his sceptical argument.

The basic dispositional theory of meaning is simply that to mean something by a word is to have a disposition to use it in a certain way. Paul Boghossian, in his discussion of rule-following, states the theory more precisely, thus:

> [Speaker] *O* means (property) *P* by predicate *S* iff (it is a counterfactual supporting generalization that) *O* is disposed to apply *S* to *P*. (Boghossian 1989, p.528. Boghossian goes on to reject this theory).

This theory, then, suggests an answer to the question as to what fact about Jones makes it the case that he means something by his words. Jones means green by 'green', according to this theory, if and only if he is disposed to apply the word 'green' to green things. Now, inasmuch as one is plainly not disposed to apply the word 'green' to every green thing that crosses one's path, and inasmuch as it is not clear just what counts as 'applying a word to

a property', this theory evidently needs some elucidation and refinement. Let us consider how advocates of arguments (1) and (2) might respond to it, however, on the assumption that it can be refined to meet these preliminary objections.

Argument (1), it will be recalled, is that there are no facts about any given speaker which choose between the different possible meaning-hypotheses which are consistent with her behaviour to date. The suggestion at hand is that facts about a speaker's linguistic dispositions do just this. Kripke's familiar criticisms of this theory are that it fails to account for the finiteness of our dispositions, and that it fails to accomodate the possibility that one might be disposed systematically to misapply a given predicate (Kripke 1982, pp.22-32). Thus, considering just the second criticism, according to this theory, if Jones is disposed to apply the predicate 'green' to a certain blue thing, then what Jones means by 'green' is *ipso facto* something which includes that blue thing in its extension. To rectify the theory, what is needed is something which determines which of Jones' dispositions to call things 'green' are erroneous. For the purpose of defending argument (1) against this basic theory, then, it is sufficient to observe that dispositional facts, by themselves, fail to determine which uses of a predicate are true and which are false, and hence underdetermine what is meant by the predicate.

There are various ways in which one might adapt the basic dispositional theory in order to try to meet this criticism. The point of rehearsing the dispositional theory and this criticism of it, however, is not to evaluate the theory's prospects as a refutation of argument (1), but to contrast this response with the

response appropriate to an exponent of argument (2). Recall that argument (2) as it applies to meaning, is that inasmuch as in doing what I take to be using a word meaningfully, I have no justification for using it one way rather than another and therefore appear not to have any way of distinguishing speaking meaningfully from merely chattering arbitrarily, I have no justification for taking there to be such a thing as speaking meaningfully. Where this argument is concerned, it appears, the dispositional theory of meaning is simply beside the point. For what in the argument is it supposed to impugn? My dispositions do not, evidently, *justify* me in going on one way rather than another. More to the point, however, I beg the question if I assume that I can unproblematically predicate of myself some dispositional property. Consider again the basic scenario: I imagine myself sitting at my window, calling certain passing objects 'green', and I wonder what justifies me in doing this as I do, and hence what distinguishes this activity from chattering arbitrarily. Clearly, exactly the same problem can be raised about my calling any given set of tendencies of mine 'the disposition to such-and-such'. Perhaps I have used the expression 'the disposition to such-and-such' in the past on many occasions, but the question is, what justifies me in calling these present activities 'the disposition to such-and-such'? If the meaningfulness of my use of the word 'green' is in question, the meaningfulness of my use of such abstruse expressions as 'the disposition to such-and-such' is unlikely to be any better off.

The basic dispositional theory of meaning, then, fails to vitiate the conclusion either of arguments (1) or (2), but it fails for very different reasons in the two cases. Where argument (1) is concerned,

facts about a speaker's dispositions are admissable as a candidate for determining what the speaker means, and must be evaluated as such. Where argument (2) is concerned, by contrast, my very ability to formulate claims about my dispositions is what is in question, whence it won't do to try to appeal to dispositional facts. The comparison to epistemology is illuminating here. The status of dispositional facts here can be compared to the status of facts about the functioning of cerebral mechanisms where epistemology is in question. Just as an exponent of argument (1) has to evaluate dispositional facts as possible meaning-constituting facts, so a naturalized epistemologist might well consider facts about the functioning of some cerebral mechanism or other to be relevant to whether $X$ knows that $p$. On the other hand, in trying to solve the problem posed by argument (2), one can no more make use of dispositional facts than can the Humean appeal to facts about the functioning of cerebral mechanisms. Someone engaged in Humean epistemology would consider appealing to any fact about the functioning of cerebral mechanisms question-begging, since knowledge of it would presuppose what is in question, namely, knowledge of continued and distinct existences generally.

The general difference between arguments (1) and (2) is illuminated by a comment of W.V. Quine's and a comparison of Colin McGinn's. In the course of discussing his rejection of the intentional and semantic idioms, Quine observes that his position is consistent with Brentano's thesis that "there is no breaking out of the intentional vocabulary by explaining its members in other terms" (Quine, 1960 p.220), which is as much as to say that (putative) intentional and semantic facts are not reducible to non-

intentional, non-semantic facts. Quine comments that where he differs from Brentano, is in his deriving from the apparent truth of this thesis the conclusion that there really are no semantic or intentional facts, where Brentano concludes that the truth of the thesis entails the need for an "autonomous science of intention" (p.221). Now, as McGinn points out (McGinn 1984, p.152), argument (1), like the argument motivating Quine, can be construed as corroborating the truth of Brentano's thesis, and the exponent of argument (1), in adopting the Reductivist Thesis that facts about meaning, if such there be, must be reducible to some other, non-intentional, non-semantic facts, can be seen as siding with Quine rather than Brentano. Without the Reductivist Thesis, however, the conclusion of argument (1) leaves open the option, perhaps unpalatable but still viable, of following Brentano's suggestion. Bracketting the Reductivist Thesis, then, what argument (1) shows is, at most, that semantic facts are not reducible to non-semantic, non-intentional facts. The conclusion of argument (2), by contrast, is that I never have any justification for taking there to be such a thing as meaning something by a word, under any circumstances. This conclusion is as much of a problem for the follower of Brentano as it is for anyone else. In short, whereas argument (1) purports to show that semantic facts are not reducible to non-semantic, non-intentional facts, argument (2) purports to show that one is never justified in taking there to be semantic facts.

## III

Argument (2), then, appears to show that I am not justified in taking there to be such a thing as following a rule or meaning

something by a word. This result is at least highly counter-intuitive, and perhaps even incoherent. It remains, then, to try to find the solution to the problem posed by this argument.

The first premise of argument (2), that in following a rule, I have no justification at any given juncture for continuing in one way as opposed to another, is, it appears, correct. In following a rule, I must always ultimately proceed from what is given —past instances of the rule, formulae, intuitions, etc.— to some action, such as writing down a number. Since, at any juncture, the course of action dictated by any of the sorts of things that can be given is always open to interpretation, the choice which I make ultimately must be made 'blindly'. If argument (2) is flawed, then, the error must be in its second premise. Recall that the second premise was that if I have no justification for carrying on in one way rather than another in doing what I take to be following a rule —that is, if I carry on 'blindly'— then I have no justification for my belief that I am following a rule and not just speaking or writing arbitrarily. This premise was the result of combining the platitude P'

(P') If I am justified in believing that I am following a rule, then, assuming I do not want to make a mistake, I am justified in believing that I am constrained at any given juncture to continue in a particular way —that I am not free to go on as I please,

with the conditional (C)

(C) If, in following a rule, I have no justification for continuing in any way, then it is not the case that I am justified in believing that I am constrained at any given juncture to continue in a particular way.

The conditional (C) was accepted on the grounds that it appeared that nothing other than what I might appeal to in advance of continuing a given rule —a formula or a feeling or a sense of

conviction, say— to justify me in going on in one way as opposed to another, can give me any reason to think that I'm constrained in how I should continue with the rule. Now, the solution to the problem is to see that one can be justified in believing that one is constrained in how one can go on, and yet nevertheless have no prior justification for continuing as one does. What makes this possible is being a member of a community of rule followers. The community makes this possible in a very mundane way: although in following a rule, I proceed 'blindly', if I go wrong, my community will let me know that I have —I will perhaps be criticized and told to proceed differently. The presence of my community's criticisms and corrections justify me in believing that I am constrained in how I can go on.

The community solution to the problem posed depends on the existence of a public practice which counts as a defeasible standard of correctness. This applies both to the rule-following problem and the parallel problem about meaning. Plausibly, I count those of my (admittedly blind) uses of words as meaningful which I can see conform to a *practice* of speaking with them —i.e., which match reasonably closely to my community's uses of the same words, and for which I am held accountable by my community. When I chatter arbitrarily, on the other hand, there is no corresponding public practice —no-one will tell me that I've chattered incorrectly. Thus, what justifies me in distinguishing my uses of the word 'green' as meaningful is that they are largely the same as the uses made of it by other members of my community, and that where our uses diverge, either I (most often) or the other members of my community (very

rarely) must revise my or their use[19]. Similarly, I am justified in distinguishing as instances of, for example, mathematical rule-following, those of my deployments of mathematical symbols which match reasonably well to the uses made of the symbols by people practiced in mathematics, and which are subject to the correction of such people.

The upshot of this is that argument (2) is unsound because its second premise is false. It does not follow, as the second premise entails it does, that because my uses of words are blind, I am not justified in believing that my uses are meaningful. I use words in public, linguistic contexts (in "language games"), in which other people use the same words as I do, and in which divergences in our uses require correction. It is these facts which justify my belief that I'm speaking meaningfully rather than chattering arbitrarily, in using a given word.

Now, in the spirit of the development of argument (2), one might press the questions, what justifies me in believing that my uses of 'green', say, are *the same* as those of my community, and what justifies me in interpreting various of their utterances and gestures as exhortations to *correct* my use? The answer to these questions is that I just *see* that our uses are the same, and I just *see* that certain utterances and gestures are exhortations to me to correct my use, much in the way that I see that there is a cup on my desk. Thus, when I say 'That automobile is green' and a peer of mine responds 'No it's not, that's not green at all', I do not *interpret* him as meaning the same as what I mean by 'green' (i.e.,

---

[19]In pointing out that it is much more rare for my community to have to revise its use than me, I mean no more than that the substantial majority is most often right.

I do not go through some ratiocinative (rule-following) process, constructing his meanings out of his non-intentionally characterized behaviour), nor do I *interpret* him to be correcting (or at least contradicting) me. That he means green and not grue or any other property by 'green' is, under the circumstances, just given. In order for me to be able to distinguish my speaking meaningfully or my following a rule from my merely chattering arbitrarily, it must be the case that I immediately (i.e., without interpreting, —to interpret is to follow a rule) perceive that certain individuals (those who share my 'form of life') sometimes have the authority to correct my use of a given word, and I theirs. The norm-governed practice of speaking a language is, so to speak, epistemically 'given'.

This characterization of the implications of argument (2) echoes John McDowell's construal of Wittgenstein's account of rule-following as a public practice, as a solution to a dilemma (McDowell 1984, esp. pp.341,2). According to McDowell, the first half of the dilemma posed by the rule-following problem is that if we construe being able to follow a rule as having an interpretation, then we are saddled with an infinite regress of interpretations ("Scylla"). The obvious solution to this, that rule-following and meaning are merely blind or unreflective, threatens to steer us to the other half of the dilemma, that rule-following and meaning are illusory insofar as they are not governed by any norms —there is nothing to talk but arbitrary, blind utterings of sounds ("Charybdis"). McDowell comments,

> The point of PI §198, and part of the point of §§201-202, is that the key to finding the indispensible middle course is the idea of a custom or practice. How can a performance both be nothing but a 'blind' reaction to a situation, not an attempt

to act on an interpretation (avoiding Scylla); and be a case of going by a rule (avoiding Charybdis)? The answer is: by belonging to a custom (PI §198), practice (PI §202) or institution (RFM VI-31). (Ibid.)

As McDowell's construal of the rule-following problem makes clear, the public practice, if it is to play a role in the solution to the problem, must be normative. The practice, that is, must be more than merely an arbitrary, uniform pattern of behaviour exhibited by some set of individuals —a mere statistical phenomenon, as it were. If it were just this, we would succumb to Charibdis. A practice is, rather, a pattern of use which counts as a (defeasible) standard of correctness. When I say of some object that it is green, I am speaking meaningfully not merely insofar as there exist other individuals who produce the sound 'green' in similar surroundings, but because a divergence between my usage and the community's would require a revision either of my use (most likely) or else a revision of the community's (very unlikely).

CHAPTER 3: A DEFENCE OF THE COMMUNITARIAN CONCEPTION OF LANGUAGE

Since *Philosophical Investigations*, or at least since Kripke's commentary on it[1], the idea is familiar that there are sceptical considerations about rule-following or meaning which lead to the conclusion that speaking a language in some sense knowable only by oneself —a 'solitary language', as I will call it— is somehow problematic[2]. Arguments which proceed in this way typically assume two burdens. First, they have to show that the sceptical considerations do entail that a solitary speaker faces the problems claimed. Second, they have to show that the sceptical considerations do not similarly undermine communal language. Now, this double burden naturally invites two related objections. First, one might argue that to the extent that the relevant sceptical considerations weigh against solitary language, they also undermine communal language. That is, one might argue that if the relevant considerations do make a sceptical case, it is a general case, and not one which selectively targets solitary language. Second, one might argue that to the extent that the resources of a community of speakers appropriately in agreement afford a solution to the sceptical considerations, so also do analogous resources that are available to a speaker in isolation. The aim of this chapter is to defend a version of the sceptical argument against these two lines of objection.

---

[1]Saul Kripke, *Wittgenstein on Rules and Private Language* (Kripke 1982).

[2]More will be said below about what is meant by 'solitary language'. Note also that the interest of the arguments I refer to is supposed to be that they undermine only solitary language, and not that they establish a general meaning scepticism.

I will begin by articulating a version of the argument in question. The conclusion of this argument, specifically, will be that one cannot be justified in believing that, in uttering given words or in doing anything else, one is speaking a solitary language. In Section II, I will respond to objections of the first kind just mentioned, and in Section III, to objections of the second kind.

I

As just indicated, the sceptical argument is directed against what I will call 'solitary language'. More precisely, it is directed against the possibility of one's being justified in believing that one is speaking a solitary language. By 'solitary language', I mean any supposed language of either of the following two types. First, any supposed language whose terms refer to things knowable only by oneself —things such as one's private sensations. Such languages, I take it, are what Wittgenstein refers to as 'private languages'[3]. And second, any supposed language contrived by a person entirely isolated from human contact. What is common to these two types of language is that, as will be seen, would-be speakers of them cannot be justified in believing that they know the meanings of their terms. In what follows, I will speak of would-be speakers of such supposed languages as 'solitary speakers'.

To set the sceptical problem, suppose first that I have been isolated since my birth —that I have never had any interaction,

---

[3]Cf. Wittgenstein: "The individual words of this language are to refer to what can only be known to the person speaking; to his immediate private sensations. So another person cannot understand the language." (Wittgenstein 1958, §243)

linguistic or otherwise, with any other person. I am, so to speak, a Robinson Crusoe since birth. Next, suppose that I have contrived what I take to be a language, and that I utter sentences of this supposed language in various situations. Consider, in particular, my uses of a given predicate of this supposed language —'green', say. For the sake of concreteness, we can suppose that 'green' in my language means, if anything, what the English predicate 'green' means. Now, suppose I say, concerning some object x, 'x is green'. Naturally, I am confident of just what it is that I am predicating of x —I know, so I think, what 'green' means. Upon reflection, however, it occurs to me to question this. I ask, first, 'What is it to know the meaning of a predicate such as 'green'?', and second, 'Given the answer to the first question, what justifies me in believing that I do know the predicate's meaning?'. These questions appear to be coherent and, prima facie, ought to be answerable. The sceptical problem arises because, as it turns out, the answer to the first question entails that the answer to the second is 'Nothing.'

Before proceeding to show why this should be so, some comments on the questions are in order. The first question, 'What is it for me to know the meaning of a predicate?', is of obvious interest. Insofar as knowing the meanings of (at least many of) the predicates and other expressions of a language is a precondition for understanding the language, and insofar as giving an account of what it is to understand a language is one of the central tasks of the philosophy of language, the task of giving an account of what it is to know the meaning of a predicate is of central importance. Moreover, it is plausible that, in many cases at least, in order to mean something by an expression —that is, in order to be able to

94

speak meaningfully in using an expression— one must know the meaning of the expression[4]. Certainly, for a solitary speaker, there wouldn't be much point in using an expression if she didn't know its meaning. In giving an account of what it is to know the meaning of a predicate, then, we would be giving a large part of an account of what it is to mean something by a predicate.

The second question, 'What justifies me in believing that I do know the meanings of the predicates of my language?' is also of interest. What concerns me, after all, is not merely the abstract question of what it is to know the meaning of a predicate, but the directly relevant question of whether I am right to think that I do know the meanings of the predicates of my solitary language. If I have discovered that to know the meaning of a predicate is to be in state X, then for me to believe that I know the meaning of a predicate is for me to believe that I am in state X. And if, for each of the predicates of my language, I have no justification for believing that I am in state X, then nor will I have any justification for believing that I know the meanings of the predicates. Of course, this would not be a cause for concern, if I could show that the belief was rational in the absense of my having any justification. That I know the meaning of a word, though, is a substantive claim —I cannot simply take for granted that I do. There are, of course, many words whose meanings I do not know, and it is

---

[4]This claim is hedged ('in many cases at least') because it is plausible that one can, in the context of a linguistic community, sometimes use an expression meaningfully without knowing its meaning. Tyler Burge's example of a speaker who mistakenly believes that he can have arthritis in his thigh (Burge, 'Individualism and the Mental', in P. French, T. Uehling, and H. Wettstein, eds., *Midwest Studies in Philosophy 4*. Minneapolis: University of Minnesota Press, 1979), plausibly shows this.

possible for me to think that I know the meaning of a word when I do not. And insofar as this is the case, I could not rest content that I am in state X if I had no justification for my belief that I am[5,6]. To illustrate this point, consider the theory which holds that to know the meaning of a predicate is for the predicate to stand in a sui generis *means that* relation to a Platonic entity of a certain sort. And suppose that the theory does not provide that the holding of this *means that* relation can be reliably inferred from facts about usage or other epistemically accessible facts. Then, on this theory, I would have no reason whatsoever to believe that my predicates stand in the *means that* relation to suitable Platonic entities, even if they did. It would not, that is, be reasonable for me to believe that I know the meanings of my predicates (even though I would know what it is to know the meaning of a predicate).

The last comment concerns the formulations of the questions. So far, what has been discussed are questions as to my *knowing the meaning of 'green'*. Now, one might wonder what the relation is

---

[5]Note that I am not presupposing here that, generally speaking, for one's belief that P to be rational, one must have a justified belief that the specific conditions which have to hold for P to be the case, hold. Thus it may be that although, for example, I do not know what conditions have to hold for it to be the case that my stomach will digest the sandwich I just ate, my belief that it will is rational all the same. Once one knows what conditions have to hold for P to be the case, however, matters change. A gastroenterologist, for example, being apprised of the conditions necessary for digestion to take place, and enquiring seriously into the question as to whether it will, would not be rational in believing that it will if, after having investigated, she had no justification for believing that the necessary conditions do hold (that the necessary stomach acids are present, etc.). I am maintaining that the philosopher enquiring into whether he knows the meaning of a predicate, is like the gastroenterologist here.

[6]This point will be taken up in more detail, in section III.

between these questions, and the parallel questions pertaining to whether I *know that 'green' means green*[7]. To begin with, it should be noted that there is an ambiguity in the latter questions. In saying that S knows that the English predicate 'green' means *green*, one might, on the one hand, mean merely that S knows a certain metalinguistic proposition to be true, viz., that the English predicate 'green' means *green*. S's knowing this metalinguistic proposition to be true is consistent with S's not knowing the meaning of 'green' —S may know this simply in virtue of knowing that all sentences of the form 'The English predicate 'X' means X', are true. In this chapter, I shall not be concerned with the questions thus construed. On the other hand, in saying that S knows that the English predicate 'green' means *green*, one may simply mean that S knows the meaning of the English predicate 'green'. On this construal, then, the parallel questions are equivalent. Insofar as this is the case, either formulations of the questions would be sufficient for this chapter. Despite this equivalence, though, I will formulate the discussion only in terms of *knowing the meaning of 'green'*, as opposed to *knowing that 'green' means green*. This is because the former formulation, unlike the latter, is unambiguous, and because the former does not presuppose, as the latter does, that in enquiring as to whether I am justified in believing that I know the meaning of 'green', I can use as well as mention the predicate. This having been said, the question as to what it is to know the meaning of a predicate such as 'green', as well as the question as

---

[7]Specifically, the parallel questions are, 'What is it to know that 'green' means *green*?', and 'What justifies me in believing that I know that 'green' means *green*?'

to whether I am justified in believing that I do know its meaning, can be taken up.

In order to begin to formulate an answer to the first question, consider how we ordinarily determine whether someone knows the meaning of a predicate. Normally, if someone consistently terms things 'green' which we —the substantial majority of speakers of English, that is— accept to be green, and makes mostly the same inferences as we would with the sentence 'x is green'[8], then we will unreflectively accept that she knows the meaning of 'green'. Conversely, other things being equal, if someone consistently terms things 'green' which we would not take to be green, or makes many inferences which we would not make with the sentence 'x is green', then we will say that she does not know the meaning of 'green'. A first suggestion, then, is this:

> Suggestion 1) Other things being equal, to know the meaning of a predicate is to be able to use it mostly as the majority of speakers of the language would.

There is an important problem with this suggestion, however. To see this, consider the case of a speaker who agrees substantially with us in what she terms 'green' and in the inferences she makes with sentences containing it, but, in those cases in which her uses diverge from ours, she persists in her divergent uses and remains indifferent to our attempts to correct her. For example, she might term things which are a deep navy blue, 'green', and remain oblivious to our objections to her use. If this person's behaviour was consistent in this respect, we would not normally say that she knows the meaning of the English predicate 'green'. Part of knowing what 'green' means is knowing that there is a *correct* use to which

---

[8]For example, she infers from 'x is green' to 'x is coloured'.

one must, like all English speakers, aim to conform. Insofar as the similarity of this person's uses of 'green' and English speakers' uses is a mere coincidence and not the result of a common aim, she does not know what the English predicate 'green' means. In a similar vein, consider the case of a machine, outfitted and programmed to term things we accept to be green, 'green', and to make suitable inferences with 'x is green' (and similarly programmed for other English expressions), but entirely unable to alter its pattern of judgements in light of experience —unable, that is, to correct itself. Such a machine would not, I submit, know the meaning of 'green', nor of any other expression. Insofar as its uses of the predicate would be simply programmed reactions, and not the result of its having aimed to speak correctly, it would not mean anything by the words it pronounces, nor would it know their meanings. What these cases show, is that to know the meaning of a predicate of a given language, one must not only use it mostly as speakers of the language would, but one must have it as one's aim to use it correctly. Suggestion (1) should be revised, then, as follows:

> Suggestion 2) Other things being equal, to know the meaning of a predicate is to be able to use it mostly as the majority of speakers of the language would, and to have it as one's aim to use it correctly.

What suggestion (2) claims, it should be emphasized, is that in order to know the meaning of an expression, one has (in part) to aim to use it *correctly*. Now, one might suppose that to aim to use an expression correctly is just to aim to use it as most speakers of the language would. It should be clear on reflection, however, that in normal speech —that is, where no propitiatory, deceitful, or other special ends are being pursued— our aim is not merely to say what the majority of speakers would say, but to speak the truth.

Thus, a person who asserts 'x is P' against doubters who assert 'x is not P', is not making a claim about what most people would say; she is making a claim about what x is. Similarly, we aim to make inferences with 'x is P' that are correct, rather than merely the inferences which most people would make. The default supposition, of course, is that the use of a predicate current among the substantial majority of speakers is the correct use, and hence that if we are in agreement with these speakers we are likely to be correct, but this need not always be the case.

Suggestion (2), then, looks like a promising answer to the first question. The problem which it poses for the solitary speaker, though, is that he will necessarily be in agreement with the majority of speakers of his solitary language (since he constitutes the whole of the population of speakers). This makes it inappropriately trivial that he knows the meaning of the predicate. In light of this, a natural third suggestion is simply

> Suggestion 3) Other things being equal, to know the meaning of a predicate is to be able to use it mostly correctly, and to have it as one's aim to use it correctly[9].

For solitary languages at least, this surely cannot be objectionable. Indeed, insofar as the default assumption, just mentioned, that the use of a predicate current among the majority

---

[9]Here and throughout, I assume that it is correct to term x 'P', if and only if x is P. I formulate my discussion in terms of what it is correct to say, because I take this to be idiomatically correct. Nothing would be lost, however, if formulations in terms of speaking truthfully were substituted throughout for my formulations in terms of correct use. Then suggestion (3), for example, would read,

> Other things being equal, to know the meaning of a predicate is to be able to use it to speak mostly truthfully, and to have it as one's aim to speak mostly truthfully in using it.

of speakers is the correct use, is right, suggestion (3) follows directly from suggestion (2)[10].

There are two qualifications to make to suggestion (3). The first qualification is occasioned by the possibility of predicates with null extensions, such as 'unicorn', 'phlogiston' or 'luminiferous ether'. Where such terms are concerned, knowing the terms' meanings cannot consist in being able (in part) to predicate them of things which are in their extensions, since there are no such things. Plausibly, in such cases, knowing the meaning of the predicate must consist entirely in aiming to make, and being able to make, correct inferences with sentences containing it. Secondly, one might object that there can be cases in which a person generally fails to use a predicate mostly correctly, and yet still knows its meaning, provided that the failures are due to mental or perceptual lapses of some kind. If it is indeed plausible that there are cases of this kind, the account of what it is to know the meaning of an expression would have to be modified to the effect that to know the meaning of an expression is to aim to use it, and to be able to use it, mostly correctly when one's mental and perceptual faculties are operating minimally reliably.

The aim of the preceding discussion, it should be emphasized, is to determine the minimal requirements which one must meet in order to know the meaning of an expression, with a view to seeing whether I can be justified, in the scenario described at the outset, in believing that I meet them. For present purposes, what is interesting about suggestion (3) is that, central to what knowing

---

[10]I will argue, in Section II that this default assumption —that the majority of speakers must be correct, most of the time— must be correct.

the meaning of an expression is, is the matter of whether one is using the expression *correctly*. That is, knowing the meaning of an expression is a matter of aiming and being able to use it in accord with a standard of use. Given this, the second part of the question —'What justifies me in believing that I know the meaning of a given expression?'— becomes, 'What justifies me in believing that I am aiming, and am able, to use a given expression correctly, i.e., in accord with the rule for its use?' In the scenario described above, in which I am solitary and I term *x* 'green', the answer is not obvious. Certainly, I have a conviction that *x* is correctly termed 'green' —after all, my unreflective disposition is to say 'x is green.' This is just to say, however, that it *seems* right to me to term *x* 'green.' In order for me to justify my conviction, what I need is some reason to believe that what seems right to me is a reliable indicator of what *is* right. Without such a reason, my belief that I am able to use the expression correctly will be mere optimism. And in order for me to be able to have such a reason, what I need, in addition to my direct epistemic access to what seems to me to be right, is some kind of independent epistemic access to what is right. That is, I need some way of judging, at least some of the time, whether a judgement of mine is right, independently of whether it seems to me to be right. Since this seems not to be something which I am able to do in the scenario described at the outset, it appears, on reflection, that I really do not know what 'green' means.

It will be recalled that 'solitary language', as I have defined it, refers not only to languages contrived by individuals in the situation I imagine myself to be in at the outset —that is, isolated

from all human contact— but also to languages whose terms refer to objects knowable only to one person, sometimes called 'private languages'. Now, insofar as members of a person's community could not, by hypothesis, know the supposed meanings of the expressions of the given person's supposed private language, they would be of no help to this person in determining whether he was able to use its expressions in accord with a standard of correct use. Thus, whatever conclusions apply to the isolated speaker will apply equally to any would-be private-language speaker, regardless of the presence or absence of other people in his environs.

The problem, then, is that the answer to the first question appears to entail that the answer to the second question —'What justifies me in believing that I do know the predicate's meaning?', is 'Nothing.' Before considering how a person who was not solitary can escape this problem, there are three prima facie plausible responses to consider.

A first response to the problem is to reject the claim that being correct is a matter of being in accord with an independent standard, and hence to reject that there is a problem about my being justified in believing that I am using my predicates mostly correctly. That is, one might maintain that Wittgenstein's slogan, "[f]or a *large* class of cases ... the meaning of a word is its use in the language" (Wittgenstein 1958, §43), is correct, on the interpretation that 'use' means *actual* use. In the case of my solitary language, then, there would be no problem about my being justified in believing that I am aiming to use, and succeeding in using, my predicates mostly correctly, since my actual uses would ipso facto be correct uses. This suggestion is motivated by the thought that, insofar as it is

I who assign the predicates of my solitary language their meanings, the correct uses cannot be independent of me. The familiar problem with this suggestion, though, is that it makes it impossible for me not to be justified in believing that I know the meaning of a predicate. The very fact of my using a predicate entails that I know its meaning, and that I am justified in believing that I know its meaning. Since the possibilities of my not knowing the meaning, and of my not being justified in believing that I know the meaning, are real possibilities which must be accomodated, this response, as it stands, cannot be right. Now, a more sophisticated variant of this position —one which avoids this problem— is to construe 'meaning is use' to mean no more than that two speakers' expressions have the same meaning if and only if they have the same use. On this construal, to say that one is conforming to a standard of use is just to say that one is using, and aiming to use, a predicate as another speaker or speakers would. As has been seen, though, this construal of the 'meaning is use' slogan is of no use to the isolated speaker, since it presupposes that there are more than one speaker of the language.

Thinking along similar lines, I might suppose, secondly, that the mere fact that I am confident about terming some things 'green' and about not so terming others, is enough to justify my conviction that I know what 'green' means. After all, it is generally the case with what I take to be predicates whose meanings I know, that I am inclined to predicate them of some objects and not others, whereas this is not generally the case with what I take to be nonsense words and predicates whose meanings I do not know (I don't go around

terming some things but not others 'glumph', say)[11]. This, in one form or another, is the basic reply of the defender of solitary language, and will be discussed in more detail in section III. The basic problem with this suggestion, though, is that my having a feeling of confidence or conviction or naturalness as I term a thing 'green' is not, by itself, evidence that it is correct to term the thing 'green'. What I need, but am not afforded by such feelings alone, is a reason to believe that the presence of the feelings as I use the term is a reliable indication that the use is correct. When I meaningfully predicate 'green' of some object, I say that the object is a member of a certain set —namely, the extension of 'green'— the constitution of which is independent of my inclinations. That my word 'green' manages appropriately to latch onto this set, or any set at all, it seems, cannot be determined by examining solely my inclinations to use the word, since my inclinations need not reflect anything more than my transient preferences.

A third response to the question is to suggest that I can be justified in believing that my uses are meaningful insofar as I can make a detached, empirical study of them and determine them to have the meanings they do. In determining that my uses conform to certain rules of use, established by an empirical science of semantics, I

---

[11]This view, it should be noted, is an epistemological counterpart to the so-called 'dispositional' account of meaning, according to which meaning something by an expression consists in being disposed to use the expression in a certain way. Since the aim here is not to solve any sceptical problems about what *constitutes* meaning something by an expression, the attempt will not be made here to evaluate the dispositional account. For a comparison of the epistemological and constitutive accounts, see Section II of Chapter 2.

would justify my belief that my uses are mostly correct[12]. To see the rejoinder to this response, notice first that the question raised about meaning can equally be posed about rule-following. Thus, instead of asking what it is for me to know the meaning of an expression, I can ask what it is for me to follow a rule, such as a rule of logic or arithmetic, or, more pertinently, a rule of scientific procedure. And, insofar as the answer is that it is to be in conformity with a standard of a certain sort, the same epistemological question arises: what justifies me in believing that I am in conformity? Thus, to answer the question 'What justifies me in believing that my uses of 'green' are mostly correct?' with the reply, 'They're in conformity with a rule of use established by empirical semantics', is merely to raise the further question 'What justifies me in believing that my taking them to be in conformity is a reliable indication that they are in conformity?'[13] In short, the objection comes down to this: if there is a problem about my being justified in believing that I am using a word as basic as 'green' in accord with a standard of use, then it is hard to see how I could begin to be justified in believing that I am follow anything as abstruse as a rule of scientific procedure, let alone how I could appeal to the outcomes of such abstruse deliberations to justify my initial belief.

If my being a member of a linguistic community is to allow me to escape this problem, then it must somehow afford me the necessary

---

[12]One line of objection to this response which will not be pursued here, is Quine's objection to the possibility of a science of semantics (Quine 1960, ch. 2).

[13]These remarks, evidently, recapitulate Wittgenstein's comments on 'rules for following rules'. See in particular Wittgenstein 1958, §§86, 139-141.

justification for believing that I am using the term 'green' mostly correctly. In a banal way, though, this is just what my community does. If I am a speaker of English and I term something that is, say, brown, 'green', then I will be told that what I have said is *wrong* —that it is incorrect to term this thing 'green'. On the other hand, if I term something which is green, 'green', my community (i.e., members of my community) will either unreflectively accept my use, or, if it happens that I am in the course of learning the meaning of the term, my community might reassure me that I have got things right. Indeed, one's accepting the authority of one's teachers as to what is and is not correctly termed 'green' is a precondition for learning the meaning of the term. Generally, my community adheres to a practice of terming certain things 'green' but not others. Insofar as it unreflectively accepts conformity to this practice, but challenges deviations from it, my community gives me justification for believing that I am correct in terming those things 'green' which I do and in not so-terming those which I do not.

The foregoing may give the impression that, in order for me to see that I am being corrected, and hence that my words are at least candidates for being meaningful, I have first to understand, and be justified in believing that I understand, the language of the community members correcting me. This would engender the problem that, in order to come to be justified in believing that one understands a language, one would have already to be justified in believing that one understands the language. My community's disagreement with, or acceptance of, my use of an expression need not, however, be communicated linguistically. All that I need be

able to recognize, in order to see that my uses are right or wrong, are the usual signs by which people express approval or disaproval. Wittgenstein describes teaching a student by means of such signs, in considering how the words 'order' and 'rule' are taught:

> In the course of teaching I shall shew him the same colours, the same lengths, the same shapes, I shall make him find them and produce them, and so on. ...
> I do it, he does it after me; and I influence him by expressions of agreement, rejection, expectation, encouragement. I let him go this way, or hold him back; and so on. (Wittgenstein 1958, §208).

This, plausibly, is how infants and people coming to new and alien cultures are taught the prevalent language.

It should also be noted that, if I am to avoid the problem of 'rules for following rules' or 'interpretations of interpretations', my seeing that someone is in agreement or disagreement with me cannot be a matter of my interpreting that person, where by 'interpreting' I mean consulting a translation manual of some sort. If it were, then the problem which my being a member of a community was supposed to solve, would simply re-emerge concerning my justification for taking myself to be following a rule in making the interpretation[14]. It must be the case, then, that normally I simply —i.e., without consulting translation manuals or drawing inferences— see that my peers agree or disagree with me.

As mentioned at the outset, the sceptical conclusion naturally invites two questions. First, it is natural to ask whether there aren't doubts, parallel to the doubts raised in the case in which I am solitary, for the case in which I am a member of a community. And second, if the communitarian answer to the initial question is an adequate answer, it is natural to ask whether a similarly 'banal'

---

[14]On this point, see Wittgenstein 1958, §§198, 201.

answer can't be formulated which does not advert to a community of speakers —perhaps one which appeals to a 'community of time-slices' of myself. These two questions reflect a general qualm about the given solution to the problem, namely, that the point at which the sceptic's doubts are rejected seems arbitrary. After all, what justifies me in not rejecting the doubts earlier, and hence resisting the need to appeal to a community, or later, which would lead me to doubt that my community affords what's needed for me to be justified in believing that I know what I mean? Robert Fogelin expresses this qualm succinctly in his discussion of Wittgenstein on private language, as follows:

> ... the public-check argument[15] relies on what I have called a general sceptical argument[16]. If this general sceptical argument shows the impossibility of all language, then its specific application to a private language is arbitrary. It is essential, therefore, to find a defense against this sceptical argument that protects a public language without at the same time being serviceable for the protection of a private language. It does not seem that this demand has been met, for when we construct what seems to be Wittgenstein's defense against a sceptical attack upon a public language, it yields a defense of a private language as a special case. (Fogelin 1987, p.183)[17]

In what follows, the attempt will be made to show that this line of objection, although appealingly even-handed, does not apply to the argument given above. In contrast to what Fogelin charges is the case for Wittgenstein's argument, the argument given above against the possibility of a solitary person's being able to justify that she is speaking a language is not what Fogelin calls a 'general'

---

[15]That is, the argument against private language which Fogelin finds in Wittgenstein.

[16]I.e., an argument "that is independent of any particular subject matter" (p.179).

[17]For remarks in much the same spirit, see Blackburn 1984, p.295.

sceptical argument. The aim below will be to show that the recommended 'communitarian solution' does indeed represent a stable position, and that it represents the only stable position.

## II

The form of the objection to be addressed in this section is this: if I cannot be justified in believing that I know the meanings of the predicates of my language in the solitary case, then nor can I be justified in believing that I know the meanings of the predicates of my language in the public case. That is, insofar as there is a sceptical argument to be made, it is an argument which undermines the public case to the same extent as the solitary. I will consider two instances of the objection.

To see the first instance of this objection, consider the following comments on 'communitarianism', due to Paul Boghossian:

> Communitarianism is a response to the perceived inability to define a distinction, at the level of the individual, between correct and incorrect dispositions [say, to use a predicate]. The suggestion that correctness consists in agreement with the dispositions of one's community is designed to meet this need. (Boghossian 1989, p.534)

As Boghossian indicates, there is a problem in specifying what the distinction between being correct and being incorrect in one's use of a word (or having a disposition to use a word correctly or incorrectly) consists in, "at the level of the individual" —a problem which arises when one is a solitary speaker as in the problem above. Very briefly, the communitarian solution to this problem is to identify being correct with being in accord with (the majority of) one's linguistic community, and being incorrect with being in disaccord. On this view, then, the totality of one's linguistic community's dispositions to use a given predicate

110

determine the predicate's extension. If the community as a whole is disposed to term $x$ 'P', then $x$ is $P$.

The objection to this communitarian account of language is that it fails to take account of the possibility of one's whole linguistic community's being mistaken. It should be clear on reflection that one's community as a whole's terming something 'P' does not ipso facto make it $P$, as this account entails it does. For example, the mere fact that the substantial majority of one's community is inclined to term a certain celestial body 'a star' is not sufficient to make it a star —it may turn out on closer inspection to be, say, a planet or moon of a planet. In this respect, the community account of language seems to be no better off than the straight-forward dispositional account which equates the extension of a person's predicate 'P' with what she is disposed to term 'P'. This latter account fails, like the community account, because it leaves no room for error where evidently there must be room for error. So, if there is a problem in defining a distinction between correct and incorrect dispositions at the level of the individual, there is likewise a problem at the level of a community[18].

As an objection to the communitarian account of language given, this is an effective criticism. The objection, however, is misdirected, insofar as the communitarian account objected to is not what is being proposed in this chapter. The problem for which the account just outlined is offered as a solution, again, is to "define a distinction" between correct and incorrect uses of a word. It is,

---

[18]For a more complete and subtle discussion and criticism of this account, see Boghossian 1989, pp.534-536.

in other words, to fill in, in non-trivial terms, the right hand side of the biconditional,

S's use of word W is correct <=> . . .

This amounts to giving a non-trivial specification of the extension of any predicate of S's language, since it is correct to term a given thing 'P' if and only if the thing is in the extension of 'P'. The role of the community in solving the problem posed in section I above, on the other hand, is *not* to determine the extensions of the predicates of the language. To see this, recall that the problem there was to determine what it is for me to know what a word of mine means, such that I'm justified in believing that I know what it means. The proposal was that to know what a word means is just to be able to use it mostly correctly. It is sufficient, then, in order for my community to furnish me with a justification for believing that I know what my words mean, that it justify me in believing that most of my uses of my words are correct. And for my community to do this, it is sufficient that I be justified in believing that its uses be merely *mostly* correct. My believing that my community is mistaken from time to time —and thus that its uses are unable to fix the predicates' extensions— is consistent with my being justified in believing that its tacit approval or open disapproval of my uses provide me with (defeasible) grounds for taking my uses to be correct or incorrect. Even community-wide uses of a predicate which are possibly erroneous can serve as a prima facie standard of correctness, provided that it is very probable that they are correct.

The response to the first objection, then, is that my linguistic community's uses of a given predicate need not be assumed to be

112

extension-determining. My community as a whole can go wrong, provided only that its uses are correct sufficiently often that they can serve as a defeasible standard of correctness. The second objection to the thesis that I am immune to the sceptical conclusion when in a community of speakers but not when solitary, concerns my justification for believing that my community's uses of its predicates are most of the time correct. If I do not have any justification for believing this, then, evidently, I will not have any justification for believing that *my* uses are mostly correct, and hence that I know what the predicates mean. Prima facie, this problem appears to be precisely parallel to the problem facing me in the scenario above, in which I am a solitary speaker. There the relevant question was, effectively, what justifies me in believing that what seems right to me is a reliable indicator of what is right? Here the question is, what justifies me in believing that what my community takes to be right, is a reliable indicator of what is right?

This question can be taken in two ways[19]. On the first way of taking it, it is assumed that my linguistic community, whom, we're supposing, agree in their uses of their words, do mean something by their words (and all mean the same thing). Then the question is, what justifies me —relative to the given assumption— in believing that what my community as a whole take to be the correct uses of their predicates, do not diverge substantially from what in fact are

---

[19]In Christopher Peacocke's terms, these are the issue of whether we are making mistakes in application of rules, and the issue of whether we are making mistakes about the existence of rules (Peacocke 1981, p.76,7). These two ways of taking the question also parallel Fogelin's sceptic's response to Wittgenstein (Fogelin 1987, p.181,2).

the correct uses? That is, what justifies me in believing that their applications of their predicates are not mostly mistaken? For example, why not suppose that, though there is very large agreement in the community as to what the correct uses of 'cubic' or 'married' or 'sweet' are, nevertheless, we are all more often than not, incorrect —e.g., most things we term 'sweet', in fact are not correctly so termed, and many things we would not term 'sweet' in fact are correctly so termed. The second way of taking the question posed above is with this assumption —that my linguistic community do mean something by their words— dropped. Then the question is, what justifies me in believing that what my community take to be the correct uses of a predicate are a reliable indicator of what in fact are the correct uses, as opposed to nothing at all? Why suppose that there is such a thing as a correct use —i.e., a meaning— as opposed to merely a conformity in dispositions to use the predicate?

The balance of the remainder of this section will be devoted to answering this question as understood in the second way. As will be seen, the answer to this constual of the question entails an answer to the first way of taking it, as well. It should be noted that this second way of taking the question is the relevant parallel to the question for the solitary case. There the issue was, what justifies me in believing that what seems right to me is a reliable indicator of what is right, as opposed to being neither right nor wrong at all? My concern was not that I might mean something different than what I appear to mean by my words, but rather that I can't justify the belief that I mean anything at all. Similarly, the concern here is to show, not that what my community take to be right doesn't diverge substantially from what in fact is right, but rather that

what my community take to be right is in fact evaluable as right or wrong.

What is being proposed is a scepticism concerning the meaningfulness of public speech, parallel to the scepticism in the solitary case. The correct response to this scepticism, I maintain, is to reject the aberrant uses of our ordinary concepts which the sceptic exploits to get his result. In this respect, the response is analogous to a familiar line of response to scepticism concerning the existence of an 'external world'. Wittgenstein summarizes this line of response in the course of discussing the concept of 'seeing'[20]:

> It [characterizing the concept of 'seeing'] is the same as when one tries to define the concept of a material object in terms of 'what is really seen'. —What we have to do is to *accept* the everyday language-game, and to note false accounts of the matter *as* false. The primitive language-game which children are taught needs no justification; attempts at justification need to be rejected. (Wittgenstein 1958, p.200)

Now, I cite this comment of Wittgentein's because it connects the response to this scepticism —rightly, I maintain— directly to the issue of whether our normal 'language-game' requires justification. Since the sceptical argument against solitary language given in section I trades on a related question of justification, it may seem that this line of response will show the argument against solitary language to be invalid, as well. Indeed, Robert Fogelin argues just this. I will respond to this in two stages. First, I will set out the response to the sceptical argument against public language, together with Fogelin's, and indicate what I take to be a flaw in

_____

[20]To be clear, Wittgenstein is objecting, not to 'external world' scepticism, but rather to a view of perception which the sceptic exploits, according to which "what is really seen" is not physical objects but sense-data.

Fogelin's argument. This will require developing the analogy between 'external world' scepticism and public language scepticism, and the response to them, in some detail. Then, in section III, I will argue that the response, properly formulated, is of no use to the solitary speaker.

Consider, to begin with, some remarks of Fogelin's. As was noted at the end of section I, Fogelin charges Wittgenstein with using a "general sceptical argument" —i.e., one which is "independent of any particular subject matter" (Fogelin 1987, p.179)— to make his case against the possibility of a private language[21]. More specifically, he says:

> In general, the sceptic exploits the distinction between seeming and being and argues that we are not in a position to decide in particular cases whether something has a characteristic or only seems to. (This is just how Wittgenstein argues against the possibility of a private language.) In everyday life this challenge does not bother us because we accept a principle of the following kind:
>
> If something seems to be *p*, then (defeasibly) it is *p*.
>
> ... If there seems to be a tree in front of me, I straightaway think there is a tree in front of me, and retreat from this belief only for good reasons. The wile of the sceptic is to reverse this presumption and demand that we anticipate and eliminate every possible circumstance that might arise and thereby defeat my presumption. (p. 180).

Fogelin's criticism of Wittgenstein is, in effect, that although in the private language case, he takes this sceptical strategy to be effective, in the public case he does not, and there is no ground for the double standard[22]. Fogelin maintains that although

---

[21]Cf. footnote 3.

[22]I do not mean to be defending the claim that the argument presented in this paper is Wittgenstein's (this would be a task for a separate paper). My point is rather merely that the present argument is not susceptible to the criticism which Fogelin levels against Wittgenstein.

Wittgenstein is right to take our 'everyday' epistemological principles to provide an adequate answer to the sceptic, he is wrong to suppose that they cannot be applied to defend the solitary speaker.

A straight-forward example of the sceptical strategy which Fogelin (rightly, I maintain) rejects, is provided by Bertrand Russell in the opening pages of his *The Problems of Philosophy* (Russell 1912). Russell's concern is, in part, the problem of the existence of an external world, which will serve as a parallel to the problem of the existence of facts about the correctness of uses of expressions of a public language. Russell begins by observing that "[i]t seems to me that I am now sitting in a chair, at a table of a certain shape, on which I see sheets of paper with writing or print" (p.1). He goes on to say, however, that "all this may be reasonably doubted" (p.1). His argument for doubting this begins with the claim that if, for example, he moves about the table he is sitting at, the patterns of light reflected off the table, and hence the table's apparent colour, will change. Likewise, he notes, the aspect which the table presents will change from being trapezoidal to rhombic and back again, as he varies the angle from which he looks at it. Generalizing this line of thought to the other senses, Russell concludes that

> ...it becomes evident that the real table, if there is one, is not the same as what we immediately experience by sight or touch or hearing. The real table, if there is one, is not *immediately* known to us at all, but must be an inference from what is immediately known. Hence, two very difficult questions at once arise; namely, (1) Is there a real table at all? (2) If so, what sort of object can it be? (p.3-4).

The first step in the sceptical argument, then, is to make the case that what we see is not what we normally take ourselves to see, but

rather some kind of perceptual intermediary —sense data, in this case— from which the presence of the object normally taken to be seen, must be inferred. The next step is to argue that neither the deliverances of the senses nor any principle of reason is sufficient to justify the required inference[23]. The conclusion is that it is not reasonable to believe in the existence of the things in question. In short, one rejects that it can simply be taken for granted that one can infer from something's seeming to be *p* to (defeasibly) its being *p*. Since what is available to the senses falls short of *p* itself, a space for doubt is opened, and it becomes reasonable to doubt whether there are real tables and chairs —an 'external world'— at all, as opposed to mere sense-data.

The analogy to the question of my justification for believing in the existence of facts about the correctness of my community's uses of the expressions of our language, is direct. In the normal course of events, I unreflectively take there to be facts about what the correct uses of expressions are. In normal situations, that is, when members of my linguistic community term a thing 'zinc' or 'a rabbit' or 'a baseball game' or whatever, what (I believe) I perceive is not just people exhibiting vocal behaviour, but rather, people making statements which are either correct or incorrect. That I perceive the utterances of my peers in this way, it should be noted, is no more controversial than that I see a desk in front of me when I enter my office. Here again, however, one might maintain that, despite the naturalness of such convictions, they can reasonably be

---

[23]In his treatment of the problem, Russell resists this step. For a criticism of Russell's strategy for evading the scepticism which his doubts lead toward, see Crispin Wright's 'Facts and Certainty', in *Proceedings of the British Academy,* 1986. pp.429-472.

118

doubted. One might argue, for example, as follows. First, it is
evident that my sense organs are each sensitive only to physical
stimuli in my environment. Second, since it is only through my
senses that I come to learn of my environment, it follows that
(paraphrasing W.V. Quine) the only facts I have to go on in
interpreting my peers are facts about the forces that I see
impinging on their sensory surfaces and their observable behaviour,
vocal and otherwise[24]. Third, since (putative) facts about the
correctness of their uses of words are not of the given kinds, it
follows that I do not learn such facts directly at all. As Russell
might have put it,

> ...it becomes evident that facts such as whether a given thing
> is correctly termed 'a rabbit', if they exist, are not the
> same as what we immediately take in by sight or touch or
> hearing. Such facts, if they exist, are not *immediately* known
> to us at all, but must be an inference from what is
> immediately known. Hence, two very difficult questions at once
> arise; namely, (1) Are there such facts at all? (2) If so,
> what sort of facts can they be?

To complete the case, it is argued that nothing available to the
senses or reason justifies the inference from what is 'immediately
known' to conclusions about the obtaining of the facts in
question[25]. Once again, then, there is a bar to inferring from what
'seems' to be the case to what (defeasibly) is the case, and a
consequent doubt as to whether there are facts of the given kind at
all.

---

[24]Cf. Quine 1960, p.28: "All the objective data [the field
linguist] has to go on are the forces that he sees impinging on the
native's surfaces and the observable behaviour, vocal and otherwise,
of the native".

[25]Quine is an example of a philosopher who argues in this way.
Note though, that his well-known argument in Chapter 2 of *Word and
Object* (Quine 1962) concerns the possibility of inferring from what
is given to conclusions about meaning. What I am taking issue with
here are the assumptions about what is given.

At least where the kind of scepticism set out above concerning
the existence of an external world is at issue, I take the general
`common sense' response to be familiar. First, we observe that
though it is in some sense possible for me to doubt that there is
at present a desk in front of me, this is not something which it is
*reasonable* to doubt. It is not clear just how a person would have
to behave to evince sincere, non-philosophical doubt as to whether
there is a desk in front of him (in a situation in which we would
all pre-philosophically take there to be one), but, supposing it to
be possible, we would not hold such a person's doubt to be
reasonable. As J. L. Austin observes (referring to a passage in
Locke[26]),

> It [the passage in Locke] suggests that when, for instance, I
> look at a chair a few yards in front of me in broad daylight,
> my view is that I have (*only*) as much certainty as I need and
> can get that there is a chair and that I see it. But in fact
> the plain man would regard doubt in such a case, not as far-
> fetched or over-refined or somehow unpractical, but as plain
> *nonsense;* he would say, quite correctly, `Well, if that's not
> seeing a chair, then *I don't know what is.*' (Austin 1962,
> p.10)

I take this point about the ordinary notion of doubt to be valid.
In ordinary situations, one simply cannot rationally doubt that what
normal people would unreflectively take to be a desk, is a desk. If
there were known to be an extremely accomplished illusionist or
prankster or designer of stage-props in the vicinity, such a doubt
might be reasonable, but such are precisely not ordinary situations.

---

[26]"the certainty of things existing *rerum natura,* when we have
the testimony of our senses for it, is not only as great as our
frame can attain to, but as our condition needs" *An Essay Concerning
Human Understanding.* Bk IV, ch. XI, section 8. Austin refers to the
citation of the passage in A.J. Ayer, *The Foundations of Empirical
Knowledge.* Macmillan, 1940. p.1.

These considerations speak against Fogelin's suggestion that the relevant epistemological principle adhered to in everyday life is 'If something seems to be *p*, then (defeasibly) it is *p'*. This principle suggests, what Locke and Russell would presumably accept, that what we see, in ordinary situations, is not simply what is the case, but rather what seems to be the case. Once this is conceded, a good part of the sceptic's case is already made. For me to allow that there merely seems to be a desk in front of me now, but to insist that I'm nevertheless entitled to infer to its actually being there, is for me to admit the correctness of the basis of the sceptic's epistemological picture. Once I have done this, resisting the sceptic's argument is more difficult than Fogelin suggests. Suppose, for example that I say, standing in my office, 'There's a desk here'. If, in reponse to the sceptic's familiar doubts, I say that my claim is based on what seems to be the case, together with my right to infer using the stated principle, then the sceptic will likely ask what justifes my use of the principle. The alternative, which Austin recommends, to having to respond or justify not responding to this last question of the sceptic's, is to reject the reasoning which leads to the supposition that the desk is not simply —that is, not mediately— seen.

Supposing Austin, rather than Fogelin, to be right here, the problem with the sceptical argument lies, not in what can or cannot be inferred from what we are taken to 'immediately experience', but rather in what the argument claims we immediately experience. In the case of Russell's argument, the step to be rejected is the transition from the claim that, seen from different perspectives, the desk will look different, to the conclusion that all that we

immediately experience is patchworks of colour, sensations of pressure apparently at our fingertips, and so forth[27]. The difficulty with this step is that it is simply false that the desk looks different, in the way Russell suggests, from different angles. The corners of my desk look *right-angled*, not obtuse or acute, from any angle that I look at it, in normal circumstances. If asked to choose which, among a carpenter's square, a similar, obtuse angled object and a similar, acute angled object, would fit the corners of my desk (or even a normal-looking rectangular desk never before examined), I would choose the carpenter's square. Similar objections apply to Russell's claims about the desk's colours. Now, to be sure, if I concentrate, I can focus my attention on the patchwork of colours which the table, seen from a certain angle, presents, and see the various quadrilateral patches as having acute and obtuse, rather than square, angles. To get the conclusion that my seeing a table is a matter of my in sòme sense experiencing such patches and then inferring to the table's presence, though, much needs to be said about the senses in which 'experience' and 'infer' are being used. For they are certainly not the senses ordinarily used: in ordinary talk, I do not say that I infer to the presence of a table standing directly in front of me, nor that I experience patches of colour[28].

---

[27]Austin's target in *Sense and Sensibilia* is A.J. Ayer, who talks about "directly perceiving" rather than "immediately experiencing", but the point is the same.

[28]Of course, one possible path to Russell's conclusion is from the fact that it is only something like this patchwork of colours which impinges on one's retinal surfaces. As W. V. Quine points out (Quine 1960, p.2), however, this attempt to defend the epistemological priority of 'sense data' itself depends on knowledge of facts about the external world —specifically, facts about the sense organs. As Quine says, "[e]ntification begins at arm's length;

122

The foregoing is plainly not, and is not meant to be, a thorough refutation of the argument briefly summarized above. For example, I have not discussed the possible rejoinder from the sceptic that his is a different, 'philosophical' type of doubt or 'seeming', than the common sense doubts and seemings adverted to by Austin. I take Austin's strategy to be to show that there is no coherent such notion of doubt or seeming, by showing that all attempts of the sceptic's to explicate it invoke expressions used in similarly abnormal ways[29]. As in the passage from Wittgenstein cited above, this line of argument implicitly takes our normal way of talking not to be in need of justification, in the sense of being demonstrably reducible to some priviledged vocabulary, such as the vocabulary of sense-data. Although I take the strategy to be effective, I do not claim to have shown here more than that it is plausible that it is.

The response to scepticism about the existence of facts as to the correctness of my community's uses of expressions is directly analogous to the response to scepticism about the existence of an external world. To begin with, we should note that the former scepticism, like the latter, contradicts what appear to be some quite basic facts of experience. For example, in the normal course of events, when someone says to me, 'Your rabbit has escaped its hutch', I unreflectively take that person to mean by her expression 'your rabbit', a certain rabbit. In the normal course of events, to

the points of condensation in the primordial conceptual scheme are things glimpsed, not glimpses."

[29]Thus, for example, see Austin's discussions (in Austin 1962) of 'directly perceived' (pp.14-19 —c.f. our 'immediately experienced'), 'looks', 'appears' and 'seems' (pp.33-43), and 'real' (pp.62-77).

doubt this, or to venture merely that *it seems* that by 'your rabbit' she meant the certain rabbit, would be almost as perverse as to doubt whether there is a desk in front of me, or to say that there merely seems to be a desk. Similarly, it seems clear that what I am shown when I am taught the meaning of a new linguistic expression is not merely in what circumstances people are likely to utter the expression, as the sceptic suggests. Rather, what I am taught is in what circumstances it is correct to use the expression. This is clear from the fact that, once I've grasped the expression's meaning, community members and I will *argue* with one another if our dispositions to use it differ, and not merely register the deviation from what's expected. Granted, the fact that the expressions of my community's language have correct uses does not typically cross my mind when I make, or when I am presented with, unproblematic mundane judgements expressed in the language. Nor, however, does it typically cross my mind when I enter my office space that there's a desk in front of me.

Once again, then, there is good reason to inquire into the sceptic's grounds for doubt. As in the case of external-world scepticism, the sceptic's argument is straight-forward. His strategy, again, is to identify what we perceive with something other than the facts in question, and to argue that the facts in question cannot be inferred from what is perceived. And, as in the case of 'external world' scepticism, this identification is a non-trivial philosophical thesis which is at odds with common sense. The sceptic's first premise, that our sense organs are each sensitive only to physical stimuli, is surely unexceptionable. What is questionable is the step from this to the conclusion that all the

objective data we have to go on in interpreting another person are the forces that we see impinging on the person's bodily surfaces and her or his observable behaviour, vocal and otherwise. Since bits of data are facts of a certain sort, such as that a meter shows a certain reading or that a person uttered certain sounds upon being aurally impacted in a certain way, this conclusion says, in effect, that the only facts we non-inferentially perceive to obtain are 'physical facts' —facts about the presence, movements, and configurations of physical objects. And this neither obviously follows from the first premise, nor is it obviously true. Concerning whether it follows, note that we have already given up, in responding to the external-world scepticism considered above, the idea that our sensory apparatus deliver to us some special class of 'immediately given' sense data from which all judgements about physical objects must ultimately be derivable. Given this, it is not clear how the limitations of our sensory apparatus can be invoked to show that we do not simply (non-inferentially) perceive such things as that someone means a certain rabbit by 'your rabbit', or that she disagrees that a certain object is correctly termed 'P'[30]. Just as the problem of the underdetermination of facts about the external world by sense-data was solved by rejecting the picture of perception which gives rise to it, then, so the problem of the underdetermination of semantic facts by physcial facts is solved by rejecting the account of perception which gives rise to it. And as

---

[30]The view being defended here, that we perceive non-inferentially the contents of others' utterances is defended by John McDowell, especially in his 'Anti-Realism and the Epistemology of Understanding', in Herman Parrett and Jacques Bouveresse (eds.) *Meaning and Understanding*. New York: De Gruyter, 1981, pp. 225-248.

to the prima facie plausibility of the conclusion, as has already been emphasized, when I make such judgements it is normally not as the result of my drawing any explicit inference —in normal cases I just see that, by 'your rabbit', my peers mean a certain rabbit.

The answer to the present question, then, as to what justifies my belief that my community's uses of the expressions of its language are evaluable as correct or not, is that I just see that they are —I neither have nor need a more explicit justification than this. The entirety of our social practice is based on our taking one another to be speaking meaningfully, and insofar as the sceptic's doubts as to the soundness of this practice are based on a dubious account of the kinds of states of affairs we can perceive to obtain, we are justified in upholding the practice. Now this view, that we normally simply see, without making any inferences, that another person is in agreement or disagreement, or that a fellow speaker means such-and-such by her words, of course is not meant to entail that we cannot be wrong. In cases in which we are wrong, though, it is no mystery how we discover that we are. Just as we find that something which we took to be a desk is not a desk, by, say, tapping it and discovering it to be made of convincingly finished cardboard, so we discover that someone whom we took to be in disagreement was in fact in agreement, by what that person says and does in later situations, and in response to further questioning. Our knowledge of whether our peers are in agreement or disagreement with us, then, is defeasible, but there is nothing mysterious in this.

It will be clear from this, what the response to the first construal of the question raised above —namely, 'What justifies me in believing that what my community as a whole take to be the

correct uses of their predicates, do not diverge substantially from what in fact are the correct uses?'— will be. Just as our having a useful practice of taking one another to be speaking meaningfully in uttering words is all the justification I need for believing that we are, our having a practice of assuming, where there is widespread agreement among speakers of the language, that a given usage is correct, is also all the justification I need for believing that the usage is correct. It can hardly be maintained that whereas we simply perceive one another's utterances to be evaluable *as* correct or incorrect, we may be drastically wrong about *when* they are correct andwhen incorrect. This would be analogous to maintaining that, whereas it is absurd for me to doubt that there is a material object in front of me, I cannot similarly dismiss the possibility that it may be a chair or a sheep, rather than the desk which it 'appears' to be.

The general response here, then, like the response to external world scepticism, is based on a rejection of the sceptic's argument that the relevant part of our normal practice or 'language game' —our talk of material objects, on the one hand, and our talk of statements being correct or true, on the other hand— must he reducible to some other, privileged vocabulary. In rejecting the sceptic's argument for the priority of the priviledged vocabulary, his argument for the need for a justification of our ordinary talk in terms of it, is undercut.

## III

In section I, it was maintained that to know the meaning of a predicate is to aim, and to be able, to use it mostly correctly, and

that a solitary speaker cannot be justified in believing that her or his uses are correct. It was also held that a speaker in a linguistic community staight-forwardly can be justified in believing that her or his uses are mostly correct. The point of section II was to respond to the objection that if the reasons are sound for thinking that the solitary speaker cannot be justified, then the same reasons will show that the speaker in a community cannot be justified, either. As was noted at the end of section I, though, one might just as well object that if there is a solution to the sceptical argument for the speaker in a linguistic community, then there is an analogous solution for the solitary speaker. This section is devoted to showing that it does not.

Simon Blackburn makes an objection close in spirit to the objection to be criticised here, in his discussion of Kripke's version of the rule-following problem. Although the problem and communitarian solution developed in section I are not Kripke's, Blackburn's objection to Kripke can be reformulated to apply to them. Paraphrasing Blackburn, then, the objection can be put as follows[31]:

---

[31]Blackburn's original objection reads as follows:
"It may be helpful to think of it like this. The members of a community stand to each other as the momentary time-slices of an individual do. So just as the original sceptic queries what it is for one person-time to be faithful to a rule adopted by a previous person-time, so the public sceptic queries what it is for one person to be faithful to the same rule as that adopted by another. Now, if the public sceptic can be by-passed by, in effect saying that this is what we do —we see each other as mutually understanding the same rule, or dignify or compliment each other as so doing, provided the exposed practice agrees well enough, then the private sceptic can be by-passed in the same way. His doubts admit of the same projective solution." (Blackburn 1984, p.294). I do not mean to impute my variant of Blackburn's objection to Blackburn himself.

It may be helpful to think of it like this. The members of
a community stand to each other as the momentary time-slices
of an individual do. So just as the original sceptic
challenges my justification for believing that my various
time-slices are using their words mostly correctly, so the
public sceptic challenges my justification for believing that
the members of my community as a whole are using their words
mostly correctly. Now, if the public sceptic can be by-passed
by, in effect saying that this is what it is for the community
as a whole to be correct —that I just *see* that such-and-such
is a correct use, in cases in which it is— then the solitary
sceptic can be by-passed in the same way. His doubts admit of
the same 'common sense' solution —I just *see* that such and
such uses of my previous time-slices are correct.

This criticism, then, takes there to be a direct analogy between the
solitary speaker and the speaker surrounded by a linguistic
community, which undermines the argument against the scepticism in
the solitary case (or, at least, the differential treatment of the
two cases)[32].

The objection can be put more precisely as follows. In the
scenario described in section I, I have before me a number of
objects which I am terming either 'green' or 'not green'. The
question posed was, in effect, what justification can I, as a
solitary speaker of my language, have for believing that it is
correct for me to term these things 'green' but not those? The
answer was that since nothing I can appeal to affords the
justification, I should not believe that it is correct or incorrect,
as the case may be. In short, I have no ground for drawing a
distinction between what seems right and what is right. The present
objection to this is that surely I *do* have a basis for drawing the
distinction, namely, the presence or absence in my 'present time-
slice' of an inclination to heed what past time-slices have said.
Thus surely I can, as a solitary speaker, term a given thing 'P',

---

[32]See also Fogelin 1987, p.183, for an objection in much the
same spirit.

take a closer look at it, then on second thought term it 'not-P'. In such circumstances, surely, it would be appropriate to say that the claim 'This is P' initially *seemed right* to me, but turned out not to be. Naturally, my latter judgement is defeasible —a yet further future time-slice of myself may, on even closer inspection, reverse my judgement yet again— but part of the point of the common sense epistemology belaboured in section II was just that such claims are always defeasible anyway. It was not because claims as to what is correctly termed 'P' are indefeasible in the public case, that the scepticism was defeated there —it was conceded that they're not indefeasible.

One might claim, in support of this objection, that the analogy here, between the individual considered as a set of momentary 'time-slices' and the community of speakers, is in all relevant respects, exact. Thus, just as a judgement of mine can disagree or accord with the judgement of another community member, so a present judgement of mine can disagree or accord with past and future judgements. And, inasmuch as the possibility of agreement and disagreement is the same in the two cases, so also is the possibility of distinguishing what is correct from what is incorrect.

The first criticism of this objection concerns the exactness of this putative analogy between a community of speakers and an individual considered as a set of time-slices. The key fact about the members of my linguistic community, it will be recalled, is that they can, by their words or gestures, simply tell me or show me that what I have said is wrong. Normally, when, as a whole, they do this, I have grounds for believing that I've made a mistake. Thus, suppose that I'm inclined, on the basis of information available to me at

a given time, to term $x$ 'P'. If I were in a community of linguistic peers, all of whom objected to my so terming $x$, I would have grounds to believe that what I had said is incorrect; if they did not object, I would have grounds for believing that it is correct[33]. Consider, however, the parallel for the solitary case. In this case, the analogy to my community's objecting to what I had said would be for my memories from earlier times —my past time-slices— to cause me to reverse my inclination. The first point of disanalogy to note is that whereas I do sometimes term things 'P' when the considerable majority of my linguistic community wouldn't (again, this is typically what is the case when I make a mistake), I would not term $x$ 'P' if my past time-slices were in near unanimity that $x$ is not correctly termed 'P'. After all, what I say on a given occasion is not completely unconditioned by what I have said in the past —if I am inclined to term $x$ 'P' now, it is because of something in my past experience of $x$ and of other things I've termed 'P'[34]. Assuming I am not speaking merely out of whimsey, I will not term something, say, 'green', unless it strikes me as similar to things I've termed 'green' in the past —i.e., which past time slices have termed 'green'.

The second point of disanalogy to note concerns the nature of the disagreement between, on the one hand, me and my linguistic peers,

---

[33]It should be emphasized, again, that cases in which one individual (perhaps a scientist or philosopher) opposes the majority and proves right, are relatively uncommon and can be accomodated as exceptions. The majority of cases, I submit, are like the case in which I term a certain glass of liquid 'a good stout', whereupon all those around me object that it is correctly termed 'a bitter', and not 'a stout'. In such cases, I should infer that it is incorrect to term the liquid 'a stout'.

[34]I am assuming that I am speaking reasonably carefully, and that any 'mistakes' I may make are not due simply to haste.

and on the other hand, me and my past time-slices. In the first case, we —me and my peers— all have direct access to $x$ as it is at present. We all come to our views on the basis of the same information. In the second case, by contrast, it is only my present time-slice which has access to $x$ as it is at present. The things termed 'P' or 'not-P' by my past time-slices were either not $x$, or were $x$ as it was at previous times (were different 'time-slices' of $x$). Thus, in the second case, it is up to me alone —my present time-slice— to determine whether $x$ as it is at present is similar in relevant respects to what my past time-slices termed 'P' or 'not-P'. If one were to make an accurate analogy in community terms for this consultation of my past time-slices, then, it would not be simply the model of ordinary correction by linguistic peers. Rather, the model would be this: when making a judgement, I listen to what each (or as many as I like) of my community members has said about things which I alone can judge to be similar or not to the thing now in front of me, and then I alone make my judgement as to whether it is correctly termed 'P'[35].

The issue, then, is whether these differences between one's being solitary and one's being in a community of linguistic peers, make

---

[35]There are other points of disanalogy besides these two. One worth mentioning concerns the fact that learning a language requires accepting the authority of certain speakers —for a child, the authority of her parents, for a non-native, that of competent speakers (and it is a familiar point that there is 'division of linguistic labour', whereby native speakers defer to the authority of 'experts' in particular fields). In the solitary case, the individual whose usage is being corrected —my present time-slice— is always the *most* experienced, and hence plausibly the most expert, among the individuals judging, since past time-slices necessarily have less experience. In the community, by contrast, the individual being corrected is characteristically *not* the most experienced, and defers to greater experience.

the difference between one's being able to justify one's belief that one is speaking meaningfully or not. The present suggestion, again, is that my basis for drawing the distinction in the solitary case between what seems right and what is right, is the presence or absence in my present time-slice of an inclination to heed what past time-slices have said. Formulated as an answer to the question posed in section I, then, the suggestion is that what justifies me, when solitary, in believing that I am aiming to use a given predicate 'P' correctly and am able to use it mostly correctly, is that in deciding whether to term a given thing 'P', I in some sense take note of, or heed, past occasions on which I termed things 'P' or 'not-P' (I consult past time-slices of myself). And my claim amounts to this: whereas the tacit acceptance or open disapproval of my peers with the judgements I make gives me reason to think that I'm conforming to an independent standard of correct use in terming things 'P', my being inclined to heed past occasions on which I used 'P', does not. Consider the relevance of the first point of disanalogy just discussed, that my past time-slices cannot, whereas my peers can, almost all be in disagreement with my present judgement. When in a linguistic community, it is always possible for me to learn, even when I have weighed as carefully as I am able to whether to term $x$ 'P', that I have likely made a mistake in terming $x$ 'P'. This is because, despite my careful deliberation, the substantial majority of my peers may disagree with me, and this will usually be good grounds for me to infer that I am mistaken. This, in turn, is sufficient reason to believe that my words' uses are subject to a standard of correctness independent of my dispositions. When I am solitary, however, this indication of the independence of

the correctness of what I say from what I am disposed to say —of what is right from what seems right— is not available to me. For me to weigh as carefully as I can whether to term $x$ 'P', when solitary, is just for me to consult all of my past time-slices, and hence to preclude the possibility of there being a substantial disagreement between what they indicate and what I have said.

The second point of disanalogy reinforces the general point. Suppose, for the moment, that there is agreement, either among my peers or my time-slices, as the case may be, that I was correct to say '$x$ is P'. Given that in the public case, all who agree with my saying this have access to $x$, it is clear that what is agreed to is that $x$ is in the extension of 'P' (we suppose, of course, that in each case we all mean the same thing by 'P'). In the solitary case, however, since my past time-slices did not have access to $x$ as it is at present, the possibility arises that I (my present time-slice) am mistaken in taking the things termed 'P' by my past time-slices to be relevantly similar to $x$ as it is at present. The possibility arises, then, that the agreement between myself and my past time-slices is attributable, not to a sameness in the object examined and an agreement as to whether it is in the extension of 'P', but to a difference in the object together with a compensating change in what the extension of 'P' is taken to be. And this point applies similarly to the case where there is *dis*agreement as to whether to term $x$ 'P'. That is, the possibility arises that the disagreement is due to a difference in the object being examined, and not to a difference as to what the extension of 'P' is taken to be. Insofar as this is the case, my being in agreement with my past time-slices as to what to term 'P' is not an indication that I am using it in

accord with a standard of correctness. The agreement can just as easily be attributed to my wavering arbitrarily in what I count as being in the extension of 'P', together with compensatory wavering as to whether to count the objects under consideration as 'the same' as objects termed 'P' by past time-slices.

The response to the objection under consideration, then, comes to this. The answer to the question 'What justifies me in believing that my uses of 'P' are mostly correct?' afforded by my linguistic peers, is clear: they tell me they're correct. The parallel answer in the solitary case, though, is inadequate. As we already know, 'It's what I'm inclined to say', will not do —as was argued in section I, this does not distinguish between my merely having a disposition to utter the predicate, and my being able to use it in accord with a standard of correctness. And it simply does not make any relevant difference to add 'It's what I'm inclined to say *having heeded my past utterances of 'P''*. That I am inclined to reflect back to past occasions of utterance in some cases but not others in making a current utterance, does not show, in the way that my peers taking issue with my current utterance does, that my utterance is evaluable as correct or not.

Now, it may still seem that a double standard is being applied here. The answer to the sceptic in the public case, after all, was an appeal to common sense: in the normal course of events, I simply perceive the expressions used by my linguistic peers to be meaningful, when they are. Surely, one might say, there can be no objection to appealing to common sense in the solitary case as well: I simply *see* that my (solitary) uses are mostly correct. As Fogelin suggests (Fogelin 1987, p.180), we should respect Wittgenstein's own

maxim here: "To use a word without a justification does not mean to use it without right" (Wittgenstein 1958, §289). Thus, although when solitary I cannot point to the difference which distinguishes my merely having a disposition to use a predicate and my having an ability to use it mostly correctly, I don't have to. It is enough that I do make a distinction.

In response to this, note first that this maxim of Wittgenstein's, although surely correct, does not help the objector[36]. For it is surely also correct that to use a word without a justification does not mean to use it *with* right. So, given that speaking with or without right —i.e., speaking meaningfully or not— is not a matter of having a justification, our question has been, how can one tell apart the two cases? To see more clearly why the reply that I simply *see* the difference, is inadequate here, whereas this was an adequate reply in section II, compare the sceptical problems in the two cases. In the community case, common sense was invoked in reply to a sceptic who raised a challenge to justify something which, in the ordinary course of events, we do not have occasion to question, viz., our conviction that we simply perceive one another to be speaking meaningfully. As was the case with `external world' scepticism, this challenge arose because the sceptic's account of what we are able to perceive entailed that we do not `directly perceive' what we normally take

---

[36]Wittgenstein's point, I take it, is that one needn't always be able to justify using a word as one does, in order for one's uses to be meaningful. When one says, for example, `This beer tastes sour', one likely will not be able to give any justification for saying this. Note, however, that being unable to justify this particular claim is not the same as being unable to justify one's claim to know the meaning of `sour'. One could easily justify this latter claim by picking out things commonly accepted to be sour (and thus showing that one is able to use `sour' mostly correctly).

136

ourselves to perceive. The common sense reply was that the sceptic, in generating his theory, uses the relevant concepts, such as 'perceive', 'see', and 'sense', in ways which diverge from their ordinary uses, with the result that it is not clear what it is that his theory is a theory of. Given this, our conviction that we often simply perceive that our peers are speaking meaningfully, was held to be unimpugned. In the solitary case, by contrast, the challenge was to justify something which we do sometimes have occasion, in the ordinary course of events, to have to justify, and which, in the public case, we know how to go about justifying. We do, in the ordinary course of events, sometimes have to justify (to ourselves or others) that we know the meaning of an expression, and we do this by comparing the way we would use the expression to the way others do, and hence showing our uses to be mostly in conformity. Thus the challenge was not a consequence of a philosophical theory, but rather of our ordinary practice —in our ordinary practice, we do have a way of drawing the distinction between knowing and not knowing the meaning of an expression[37]. What distinguishes the solitary case from the community case, is that there is no ordinary practice to appeal to. It simply is not a common practice for people to invent solitary languages, at least not languages in the normal sense, with words which can be spoken or written and compositional rules. A 'solitary language' is not, in the way that a public

---

[37]Cf. Austin's comment, made in the course of discussing the word 'real':
"Of course ... we make a distinction between 'a real x' and 'not a real x' only if there is a way of telling the difference between what is a real x and what is not. A distinction which we are not in fact able to draw is —to put it politely— not worth making" (Austin 1962, p.77).

language is, an ordinary, unproblematic object of common experience. In the solitary case, then, the flat-footed reply that I simply *see* that my uses are mostly correct, is not, as it is in the public case, simply an appeal to what we normally take to be the case. For this reason, some account has to be given of why the solitary case is analogous in relevant ways to the community case. And, as I have tried to show, this undertaking is unsuccessful.

BIBLIOGRAPHY

Austin, J.L. (1962) *Sense and Sensibilia*. Oxford: Oxford Press.

Blackburn, Simon. (1984) 'The Individual Strikes Back'. *Synthese*, Vol. 58, pp.281-301.

Boghossian, Paul (1989). 'The Rule Following Considerations'. *Mind*, Vol. 98. pp.507-549

Cummins, Robert (1975). 'Functional Analysis,' *Journal of Philosophy*, Vol. LXXII. pp.741-765.

Dretske, Fred (1986). 'Misrepresentation'. Reprinted in Lycan 1990, pp.129-143.

Dretske, Fred (1988). *Explaining Behaviour: Reasons in a World of Causes*. Cambridge, Mass.: MIT Press.

Dretske, Fred (1990). 'Précis of "Explaining Behaviour: Reasons in a World of Causes",' *Philosophical and Phenomenological Research*. Vol. 1, No. 4. pp.783-786.

Fodor, Jerry (1984a). 'Psychosemantics'. Reprinted in Lycan 1990, pp.312-337.

Fodor, Jerry (1984b). 'Semantics, Wisconsin Style', *Synthese*, Vol. 59. pp.231-250.

Fodor, Jerry (1987). *Psychosemantics*. Cambridge, Mass.: MIT Press.

Fodor, Jerry (1990). *A Theory of Content and Other Essays*. Cambridge, Mass.: MIT Press.

Fodor, Jerry (1990a). 'A Theory of Content, I: The Problem', in Fodor 1990, pp.51-87.

Fodor, Jerry (1990b). 'A Theory of Content, II: The Theory' in Fodor 1990, pp.89-136.

Fogelin, Robert. (1987) *Wittgenstein*. 2nd ed.. New York: Routledge.

Goodman, Nelson (1983). *Fact, Fiction, and Forecast*. Cambridge, Mass.: Harvard Press.

Grice, Paul (1957) 'Meaning'. Reprinted in Paul Grice, *Studies in the Way of Words*. Cambridge, Mass.: Harvard Press. pp.213-233

Horwich, Paul (1990). 'Wittgenstein and Kripke on the Nature of Meaning', *Mind and Language*. Vol. 5, No. 2. pp.105-121.

Hume, David (1739). *A Treatise of Human Nature*. Oxford: Oxford Press, 1978.

Kripke, Saul. (1982) *Wittgenstein on Rules and Private Language.* Cambridge, MA.: Harvard Press.

Lettvin, J.Y., et. al. (1959) 'What the Frog's Eye Tell's the Frog's Brain', *Proceedings of the IRE,* Nov. 1959. pp.1940-51.

Lycan, William (1990). *Mind and Cognition.* Oxford: Basil Blackwell, 1990.

McDowell, John (1984). 'Wittgenstein on Following a Rule'. *Synthese,* Vol 58. pp. 325-363

McGinn, Colin (1984). *Wittgenstein on Meaning.* Oxford: Basil Blackwell.

Millikan, Ruth Garrett. (1984) *Language, Thought and Other Biological Categories.* Cambridge, Mass.: MIT press.

Millikan, Ruth Garrett. (1990a) 'Seismograph Readings for "Explaining Behaviour"', *Philosophical and Phenomenological Research.* Vol. 1, No. 4. pp.807-812.

Millikan, Ruth Garrett. (1989) 'In Defense of Proper Functions', *Philosophy of Science,* Vol. 56. pp.288-302.

Millikan, Ruth Garrett. (1990b) 'Truth Rules, Hoverflies, and the Kripke-Wittgenstein Paradox' *The Philosophical Review.* Vol. XCIX, No. 3 pp.323-353.

Peacocke, Christopher. (1981) 'Reply: Rule-Following: The Nature of Wittgenstein's Arguments'. in S. Holtzman and C. Leich, eds. *Wittgenstein: to Follow a Rule.* London: Routledge and Kegan Paul.

Quine, W.V.O. (1960). *Word and Object.* Cambridge, Mass.: MIT Press.

Quine, W.V.O. (1969). 'Epistemology Naturalized', in *Ontological Relativity and Other Essays.* New York: Columbia Press, 1969.

Russell, Bertrand. (1912) *The Problems of Philosophy.* Oxford: Oxford University Press.

Wittgenstein, Ludwig (1958). *Philosophical Investigations.* Oxford: Basil Blackwell.

Wittgenstein, Ludwig (1978). *Remarks on the Foundations of Mathematics.* Cambridge, Mass.: MIT Press.