SEMANTIC NORMS AND TEMPORAL EXTERNALISM

by

Henry James Duncan Jackman

BA, Columbia, 1988

MA, Pittsburgh, 1991

Submitted to the Graduate Faculty of

Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

1996

SEMANTIC NORMS AND TEMPORAL EXTERNALISM

Henry Jackman, PhD

University of Pittsburgh, 1996

We typically understand an individual's thoughts and utterances in a way that ties their contents to the make-up of her physical environment and the linguistic usage of her community. This practice has frequently been taken to be in tension with the intuition that the content of one's thoughts and utterances must ultimately be explained in terms of facts about one's *own* attitudes and behavior. This perceived tension is manifested in cases where the individual's own beliefs and usage purportedly underdetermine or even misidentify what she is standardly treated as referring to by her terms. I argue that the individual's beliefs only seem to underdetermine or misidentify the referents of her terms in these cases if one presupposes a comparatively impoverished conception of what her beliefs are. The beliefs an individual speaker associates with a given term extend far beyond the handful of sentences she would produce if asked to list such beliefs. Speakers have an implicit, but rich, understanding of their language as a shared temporally extended practices about which they can be mistaken. Once this implicit understanding of language is factored in, our practice of tying what an individual means to her physical and social environment turns out to be justified by consistency requirements upon the individual's own beliefs.

Indeed, our implicit understanding of language justifies more than merely tying a speaker's thoughts and utterances to her social and physical environment. Our implicit understanding of languages as temporally extended practices turns out to justify our important, but seldom noticed, habit of reading present conceptual developments back into the thoughts and utterances of our past selves and our ancestors. We can both endorse a picture of linguistic norms that is methodologically individualistic and allow that future use (as with communal use) contributes to what we mean by our terms. External factors are relevant to what we mean because we implicitly take them to be so, and our practice of incorporating physical, social and temporal factors into our understanding of others reflects our often deep commitment to a picture of language as a shared practice extending through time.

Preface

I've profited from conversations with many people while writing this dissertation, in particular I'd like to thank Dave Beiseker, Akeel Bilgrami, Jim Conant, Dan Everett, Richard Gale, Mitch Green, Mark Mcculaugh, and Ram Neta.

Most of all, I'd like to thank Robert Brandom, Joe Camp and John McDowell for their advice and encouragement over the past six years. My weekly meetings with Robert Brandom not only provided me with valuable insights and advice at almost every stage of the writing process, but also prevented me from falling into any of those prolonged fallow periods distressingly characteristic of graduate life. While Joe Camp remained the most elusive member of my committee, his pointed questions about various chapter drafts continually forced me to rethink the central presuppositions of the project, and the dissertation would be considerably worse without them. In addition to improving the style, spelling and punctuation of virtually every sentence in the dissertation, John Mcdowell's generous and patient comments have greatly contributed to my understanding of all of the issued discussed, and others besides.

Finally, I'd like to thank the Canada Council, whose generous support allowed me to complete this dissertation considerably more quickly than I would have otherwise.

TABLE OF CONTENTS

Introduction

There has frequently been taken to be a tension, if not an incompatibility, between "externalist" theories of content (which allow the make-up of one's physical environment and the linguistic usage of one's community to contribute to the contents of one's thoughts and utterances) and the "methodologically individualist" intuition that whatever contributes to the content of one's thoughts and utterances must ultimately be grounded in facts about one's own attitudes and behavior. In this dissertation I argue that one can underwrite such externalist theories within a methodologically individualistic framework by understanding semantic norms in terms of the need to reach, for each of one's terms, a type of "equilibrium." Each speaker's commitment to making her *own* beliefs and applications consistent allows one to incorporate these 'external' factors into the contents of their thoughts and utterances in a way that remains methodologically individualistic.

Methodologically individualistic accounts are typically taken to be unable to incorporate 'external' factors such as the world's physical make-up or communal usage because of arguments suggesting that the individual's own beliefs and usage underdetermine or even misidentify what, according to externalist accounts, they mean by their terms. These arguments, however, only seem plausible if one presupposes a comparatively impoverished conception of the individual's beliefs. The beliefs a speaker associates with a given term extend far beyond the handful of sentences they would produce if asked to list such beliefs. In particular, speakers have an implicit, but rich, understanding of their language, their world, and the relation between them. Speakers typically understand languages as shared temporally extended practices about which they can be, both individually and collectively, mistaken. Once this conception of language is taken into account, the ascriptions which purportedly forced 'non-individualistic' conceptions of content upon us (particularly ascriptions which seemed to tie what we meant to social use

rather than our own beliefs) turn out to be ultimately grounded in the individual's own beliefs.

Indeed, our self-conception does much more than merely underwrite 'non-individualistic' ascriptions. Our understanding of languages as temporally extended particulars turns out to underwrite our important, but seldom noticed, practice of retroactively reading present conceptual developments back into the thoughts and utterances of our past selves and ancestors. We can both endorse a picture of linguistic norms that is, ultimately, methodologically individualistic and allow that future use (as with communal use) contributes to what we mean by our terms. External factors are relevant to what we mean because we take them to be so, and the ascriptions we make reflect our often deep commitment to a picture of languages as shared practices extending through time.

The course of the dissertation runs roughly as follows. The first chapter argues that the possibility of misrepresentation can be understood as arising from the tension between those items to which we apply our terms (their putative extensions) and the beliefs we associate with them (their general characterizations). Both information-based and inferential role semantics are presented as unsuccessful attempts to explain semantic norms in terms of just one of these two elements. It is further argued that a term's actual extension should be understood as what one gets when one reaches an "equilibrium" between a term's general characterizations and its putative extension. However, when semantic norms are understood this way, a number of problems become salient. There is no guarantee that there will an equilibrium to be found for the practices associated with a term, and even if one can be found, there is no guarantee that there will be *just* one. Bifurcationism and indeterminacy about meaning are presented as the outcomes of these two possibilities respectively.

Discovering whether (1) what a particular term means remains indeterminate until an equilibrium is actually reached for it, or (2) whether reaching an equilibrium makes

manifest what the term meant all along, is one of the overarching projects of the dissertation. Doing so, however, requires finding an equilibrium for the interplay between the putative extensions and general characterizations of semantic terms such as "meaning" and "reference," and many contemporary disputes about meaning and content can be understood as the product of conflicting attempts to find such equilibria. The first chapter presents a number of general beliefs about meaning along with a number of ascription types brought to our attention by Kripke, Putnam, Burge and Wilson that provide evidence for the putative extensions of our semantic terms. The next three chapters investigate just how much of the putative extension made manifest in these ascriptions can be incorporated into an equilibrium for our semantic terms.

The second chapter argues that the "expressive constraint," which requires that speakers have beliefs with the same content as their sincere utterances, effectively forces whatever equilibrium we find for either thought or utterance content to be an equilibrium for both. Given that this 'internal' connection between thought and utterance content required by the expressive constraint makes finding an equilibrium for our semantic terms considerably harder, one might try to do without it. However, attempts to explain away the constraint as an overstatement of a merely *empirical* generalization about the frequent correlation between thought and utterance content are discussed and found to face serious difficulties. Not only do most philosophical accounts of thought and language commit themselves to the expressive constraint, but a number of counterintuitive consequences also follow from giving it up. In particular, rejecting the constraint makes problematic the intelligibility of successful communication and belief attribution. Indeed, it is argued that if the constraint is given up, not only would there be little reason to think that even our *self-*ascriptions would usually be true, but there would also be nothing wrong with the sincere utterance of a 'Moore sentence' such as "Carburetors are expensive, but I don't believe it" (since one could have good reason to think that one lacked beliefs with the content of even

one's sincere utterances). After establishing that we should not give up the expressive constraint, the chapter argues that the two most obvious explanations of the constraint (giving an independent account of linguistic meaning and understanding thought content in terms of it, and giving an independent account of thought content and understanding linguistic meaning in terms of it) are both inadequate. The chapter ends with a less reductive and more synthetic account of the constraint that stresses the *interdependence* of belief content and linguistic meaning stemming from the fact that our language is itself one of the things about which we have many beliefs.

The third chapter picks up on some of the themes introduced in chapter two, and argues that we should incorporate the non-individualistic ascriptions that Burge appeals to into an equilibrium for our semantic terms. Not only are such ascriptional practices themselves fairly deeply entrenched, but they also cannot be clearly isolated from the extremely well entrenched practices that Kripke appeals to. Attempts to dismiss or 'explain away' non-individualistic ascriptions by appealing to more 'pragmatic' factors are shown to rely upon a number of unrealistic assumptions about what we take ourselves to be doing when we make ascriptions or defer to correction. Even those ascriptional practices that might have seemed to favor individualism (those associated with slips, malapropisms and serious misunderstandings) turn out to be best understood within an non-individualistic framework.

General beliefs associated with our speaking a shared language whose terms have determinate extensions favor incorporating the non-individualistic ascriptions, and the third chapter argues that the most frequently cited candidates for entrenched general beliefs incompatible with non-individualistic ascriptions (those relating to self-knowledge and behavioral explanation) turn out to be compatible with them after all. Indeed, accounting for self-knowledge ultimately turns out to cause more problems for the individualist. Reconciling our ascriptional practices with these general beliefs will involve stressing (1)

that the proposed equilibrium is *non*-individualistic rather than *anti*-individualistic, (2) that the *de dicto*/*de re* distinction is a distinction between types of ascriptions, not between types of beliefs, (3) that the position is *methodologically,* rather than *ascriptionally*, individualistic, and (4) that the position relies heavily on our *self*-interpretations to fund the non-individualistic ascriptions.

As stated above, the third chapter stresses the importance of our self-interpretations when reaching an equilibrium for our semantic terms, and argues that our practice of interpreting others in terms of social usage should be incorporated into an equilibrium because speakers understand *themselves* as engaging in a shared practice when using language. In much the same way, the fourth chapter argues that, because we understand this shared practice as itself extending through time, we can adopt a "T-externalistic" position in which we endorse the everyday ascriptions of content that reflect a sensitivity to future usage. Most of the potentially unintuitive consequences of allowing future use to contribute to what we mean (such as, say, those associated with self-knowledge and behavioral explanation)   will have already been defused if we have succeeded in incorporating the non-individualistic ascriptions Burge appeals to. Furthermore, the ways in which those consequences were made palatable for the non-individualist discussed in chapter three work equally well for the T-externalist. Alternative ways of describing the relation between present and future usage are discussed and found to have a number of unintuitive consequences of their own. Finally, it is argued that since attempts to find a more temporally-bound equilibrium turn out to be no more intuitive than their T-externalistic competitors, there is no reason why we cannot endorse our temporally sensitive ascriptions in these cases.

1. Putative extensions and general characterizations

I have called many things "cats," my parents had three, my neighbor has two, I see them constantly in pet stores, on the street, etc. I also have a number of beliefs about cats: they have a certain appearance, they are animals, they are part of a biological kind closely related to lions and tigers, they like milk, they don't like getting wet, they can climb trees and fall great distances, etc. This is not unusual. For most[1] terms in my language there is a range of cases in which I would be willing to apply them, and a number of (at least implicit) beliefs I have about the type of thing the terms are supposed to pick out.[2] I'm not unusual in this respect. Each of us has a number of beliefs associated with a given term and is disposed to apply it in a range of cases (though these beliefs and applications may vary from person to person). We can call the conditions that prompt the use of a given term its "putative extension,"[3] and those more general beliefs about what one is talking about the "general characterization" associated with that term.[4]

The general characterizations will thus include all of the sentences we would endorse that include the word in question. They are, very roughly, our beliefs about or pertaining to the types of things that fall within the putative extension. For instance, when we refer to things as "chairs," we may have a number of beliefs about them (that they are pieces of furniture, people sit on them, they have backs, they have legs, etc.), and these characterizations are not independent of the putative extension. Indeed, they are often a product of our encounters with (and occasional reflections upon) the things that fall within the putative extension. Furthermore, while one can include all of the agent's beliefs in the

---

[1] I leave open the possibility that a different story may need to be told for connectives, etc.

[2] For other discussions of these two elements in a words use, see Brandom 1994 (especially pp. 119, 129, 427), Dummett 1973, and Sellars 1954.

[3] We may want to include something like the term's perceived anti-extension under the general heading of the putative extension (i.e.: not just the things we are sure are x's but also those things we are sure are not x's).

[4] It should also be noted that the general characterizations may just be presuppositions rather than explicitly held beliefs, and they may also be fairly undefined. (See appendix F.)

general characterization, there will be some privileged sub-set of these characterizations that will be considered less negotiable than the others.

For instance, for an "observational" term such as, say, "pen", one discovers a term's putative extension by discovering the stimuli the presence of which prompt the term's application. One discovers the term's general characterization by discovering what 'inferential transitions' its user is disposed to make with it. The former is manifested in a speaker's ability to pick out, say, a pen when asked to fetch one. The latter is manifested in his ability to describe pens, his belief that one can write with pens, that ball-points and rollerballs are both pens, etc. Both of these sets of abilities are important parts of a speaker's linguistic competence, and while the two usually go together,[5] they seem to be psychologically distinct.[6]

The putative extension of an oberservational term like "pen" might, then, bear some resemblance to Quine's notion of "stimulus meaning," which Quine defines as our "total battery of present dispositions to be prompted to assent or to dissent from the sentence."[7] Quine believes that the stimulus meaning of a given term can do more or less justice to its 'real' meaning. For an observation term like "red" the stimulus meaning may do "full justice" to its meaning, while for a term like "bachelor" (for that the speaker's use relies heavily on "collateral information")[8] it does not. In much the same way, we might say that the putative extension of a given term can be more or less independent of the general beliefs. In the case of theoretical terms, the putative extension may be tied quite closely to, and even dependent upon, aspects of the general characterization, and this may also be true of the more 'inferential' applications of observational terms: there are situations where I

---

[5] But not always; one could learn to identify pens very well without knowing what purpose they served, or know quite a bit about their function and history without necessarily being able to recognize them.

[6] For instance, one can have one set of abilities impaired without severely damaging the other. (See Marconi 1995 (esp. pp. 141-3) for descriptions of cases where speakers have lost much of their ability to identify objects while their ability to associate names with descriptions or synonyms remained unimpaired.)

[7] Quine 1960, 39. Though, technically, the putative extensions bear more of a resemblance to Davidson's modification of stimulus meaning, with the 'stimulus' being distal rather than proximal. (See Davidson 1990.)

[8] Quine 1960, 42.

may say a cat is present just because of general beliefs I have about the noises they make, the types of tracks they leave, etc.

It should be noted that putative extensions and general characterizations each have semantic programs associated with them. "Inferential-role semantics" focuses on the general characterizations associated with a term, while "information-based semantics" focuses on the term's putative extension. That is to say, these programs, while perhaps not *identifying* meaning with one of the two aspects of our linguistic practice, still treats it as *determined by* the relevant aspect (for a discussion of this, see section 3.2).

.

2. Tension between the two

There is thus a range of cases in which we apply a given word and certain general beliefs we have about the type of thing the word applies to. These two come into conflict when we discover that there are items to which we apply a given term but that do not correspond to our general conception of the type of thing that term should apply to.[9]   In such cases we can either stop applying the term to the object in question or we can alter our general conception of the type of thing the word applies to.   An instance of the first alternative can be seen in the development of the use of "fish."   We have a set of general beliefs about fish such as: they live in the water, they have gills rather than lungs, are cold blooded, have scales, don't suckle their young, etc. The fact that whales don't satisfy the last four of these descriptions caused at least some of us to change the putative extension of "fish."[10] On the other hand, if the general characterizations are too far out of line with the putative extension, they will usually be revised or given up.   For instance, we would probably give up the claim that chairs have legs rather than claim that all of the modern legless chairs are not 'really' chairs. In some cases, the entire framework of the *type* of

---

[9] Or vice versa. It should also be noted that there may be conflicts within these two factors themselves. We may apply our terms in an irregular manner (sometimes applying the term to an object, sometimes not) and our general ideas about the type of thing the term applies to may be inconsistent.

[10] This description is perhaps a little misleading, since it is given at the social level, and the change in usage of an entire society will take longer and not proceed in such a simple fashion.

thing one is talking about is thrown out when it turns out to be in conflict with too much of the putative extension.

A fairly clear instance of this is our eventually giving up the assumptions that trees or lilies form any type of 'natural kind.' Deciduous trees and evergreens bear no close taxonomic relation to each other, and many lilies are more closely related to onions than they are to other lillies, but we don't say that evergreen's aren't 'really' treees or that onions are really a type of lilly. The same applies to "reptile" as well (crocodiles being cladistic sisters of birds, not lizards). Of course it is worth noting that one can take an extremely hard line, and keep with the natural kind intuition in these cases and just dump the whole practice ("[The] cladogram of "higher" primates contains an interesting implication: there is no such thing as an ape, at least as usually defined").[11]

Once such a basic belief is given up, the type of belief the general characterizations must focus on will change. On perhaps an even more basic level, changes in what we take to be involved in being, say, a natural kind can reverberate throughout the entire system. For instance, if we change our conception of what a biological kind is (possibly because of our previously having dealt with an isolated range of cases) we may have to reconsider the extensions of many other terms.

We may say that an aspect of our usage is "entrenched" to the extent that it is difficult to dislodge when it comes into conflict with other aspects of the term's usage, and which element of a speaker's usage is given up is a function of the comparative entrenchment of these elements. Aspects of our usage are frequently entrenched simply by being long-standing, but many other factors can entrench a belief or application. The extent to which it is supported by or supports other entrenched beliefs (including one's theoretical or ideological commitments) could, for instance, contribute to the comparative entrenchment of some aspect of a term's use.

---

[11] Gould 1983, 358.

Still, entrenchment is, to a large extent, a function of numerical support. Fifty beliefs or applications will, collectively, tend to be more deeply entrenched than any single belief or application. For the most part, if a small percentage of the cases that fall within the putative extension conflict with the general characterization, then those cases will tend to be removed from the putative extension, while if a large percentage of them do, it will be the generalized characterization that is modified. Of course, deciding what to give up will not always be a simple matter of finding out what percentage of the items in question satisfy a particular general characterization. Some items that a term is applied to (and some aspects of its characterization) may be more deeply entrenched than others. For instance, some of the examples of what I call "chair" are more central (paradigmatic/prototypical) than others. If our general characterization of chairs ruled out the bean-bag chairs in my parents' basement, I might be willing to say that beanbags were not 'really' chairs. On the other hand, if the characterization ruled out the formal chairs in my parents' dining room, I would probably say that there was something wrong with the characterization.[12] In much the same way, some of the general beliefs about chairs may seem more central than others: that chairs must have legs is something that can be (and probably has been) given up, while that we can sit in chairs may be less negotiable.[13] The comparative entrenchment of various elements within a term's putative extension and general characterization may vary from person to person, and it may be, at least to a certain extent, indeterminate. Which elements we will be willing to give up may depend upon the order with which we are faced with various conflicts. (The significance of this will be discussed in chapter four.) When there are no conflicting elements left in a terms usage, that term's usage can be said to be at an "equilibrium," and a terms 'real' extension can be identified with the putative extension it would have at such an equilibrium. (This will be discussed in greater detail below.)

---

[12] And this would not change even if bean bags were, in fact, more common than formal chairs.

[13] Of course, it may turn out that there are no strictly necessary and sufficient conditions of being a chair, in which case one would look for some larger set of conditions most of which would have to be satisfied for something to be a chair. (See Rosche 1975, Jackendoff 1988.)

Once this picture is in place, it should be clear that there will be a number of intermediate cases where it is not clear which should be modified, the general characterization or the putative extension. Indeed, the whale/fish case is one where different speakers can have different reactions as to which element should be modified.

For example, Linnaeus[14] may have initially put whales within the putative extension of "fish." However, he also believed that fish and mammals were mutually exclusive natural kinds. On investigating what distinguished these kinds, he discovered that, while mammals had lungs and warm blood, fish had gills and cold blood. These beliefs about what distinguished fish from mammals were incompatible with the inclusion of whales within the putative extension of "fish," and since these beliefs were more deeply entrenched than the application of the term to whales, Linnaeus came to understand that application as mistaken. He came to the realization that whales were not (and never had been) fish.

On the other hand, if the disposition to call whales "fish," or the belief that "fish" picks out a *functional* kind, is more deeply entrenched than the belief that "fish" and "mammal" pick out mutually exclusive natural kinds, one will characterize the latter belief as mistaken. Such a stance is taken by Melville's Ishmael:

> In his System of Nature, AD. 1766, Linnaeus declares, 'I hereby separate the whales from the fish.' But of my own knowledge, I know that down to the year 1850, sharks and shad, alewives and herring, against Linnaeus's express edict, were still found dividing the possession of the same seas with the Leviathan. The grounds upon which Linnaeus would fain to have banished the whales from the waters, he states as follows: 'On account of their warm bilocular heart, their lungs, their movable eyelids, their hollow ears, penem intrantem feminam mammis lactantem' and finally 'ex lege naturae jure meritoque' I submitted all this to my friends Simeon Macey and Charley Coffin, of Nantucket, both messmates of mine in a certain voyage, and they united in the opinion that the reasons set forth were altogether insufficient. Charley profanely hinted they were humbug.... I take the good old fashioned ground that the whale is a fish, and call upon holy Jonah to back me. This fundamental thing settled, the next point is, in what internal respect does the whale differ from other fish. Above, Linnaeus has given you those items. But in brief, they are these: lungs and warm blood; whereas, all other fish are lungless and cold blooded.[15]

Different interests on the part of various speakers will make different aspects of the general characterizations and putative extensions seem more central than others, fishermen and

---

[14] Swedish botanist (1707-78). This reconstruction of his thought process has no pretensions to historical accuracy.
[15] Melville, 1983, 936.

scientists being interested in 'functional' and 'biological' classification respectively.[16] These differences in what they hold to be central will lead fishermen and scientists to accept different ways of resolving conflicts between aspects of the term's use. If the parties involved have such 'incommensurable' commitments, there may be no way to settle disputes about whether, say, it is really correct to call whales "fish."[17] In such cases, it may be better to say that the conflicting parties simply refer to different things by the term.

Fairly notorious instances of unresolved tension between the general characterization and the putative extension can be seen in cases such as our use of philosophically controversial terms such as "knowledge," where some accept general characterizations that entail that we don't really know anything (or know next to nothing), while others see this conclusion as grounds for reevaluating the general characterization and replacing it with a new one that is more in keeping with the putative extension.[18] In these more deeply contested cases, it can seem particularly difficult to come up with a definitive answer as to how the tension is to be resolved. Certainly, one will not expect there to be anything like an algorithm that would tell one which of the two competing claims will have to be given up.[19]

Still, this will not always be the case in such disputes. In particular, if there is no equilibrium available corresponding to one party's use, that party is probably best understood as simply mistaken. As a result, one may argue that, say, Ptolemaic astronomers meant what we do by "planet," "star" etc. because, while they had a number

---

[16] Mapping one's terms on to natural kinds and mapping one's terms on to functional kinds are both perfectly intelligible and frequently implementable projects, and one might have initially thought that one could do both with the same terms.

[17] Indeed, even if we were, as Kripke and Putnam suggest, initially committed to treating terms like "fish" as natural kind terms, it has, of course, turned out that fish are not, on at least some accounts, a biological kind at all (lungfish being cladistic sisters with land animals, not the other fish). One should also note that Ishmael appeals to the testimony of friends from *Nantucket* (then the preeminent whaling town), since the appeal makes it clear that he is not simply being idiosyncratic, but is rather aligning himself with a different expert community (which includes, of course, "holy Jonah").

[18] Such terms may stand for what are, in W. B. Gallie's terminology, "essentially contested concepts." (See Gallie 1956.)

[19] Nor should one expect that such issues to be settled by appealing to some causal fact that would settle what was really referred to (for a discussion of such theories, see appendix A).

of differently entrenched beliefs, there is no equilibrium governing those beliefs that differs from the equilibrium governing ours.

3. How both are central to meaning

While both the putative extensions and general characterizations play an important part in determining the meaning of our words, philosophical accounts of meaning often overemphasize one of the two. In particular both contemporary "inferential-role" and "information-based" semantics are unable to achieve a robust conception of objectivity because they focus upon one of the two factors and miss out on the dynamic relation between them.

3.1 Normativity

An excessive focus on either a term's putative extension or its general characterization inevitably deprives our use of language of aspects of the normativity it (manifestly) seems to have. If the meaning of a term, "X," is determined solely by its general characterization,[20] then it seems impossible for us to be wrong about the types of things Xs are. If we come to believe that, say, all reptiles have forked tongues, then that belief simply becomes part of what we mean by "reptile" and hence (analytically) true (though, of course, it could turn out that there are no reptiles).[21] As a result of this, it could turn out

---

[20] A fairly clear instance of the overemphasis on the general characterizations can be seen in Mill's theory of general names:

> All concrete general names are connotative. The word *man*, for example, denotes Peter, Jane, John, and an indefinite number of other individuals, of whom, taken as a class, it is the name. But it is applied to them, because they possess, and to signify that they possess, certain attributes. These seem to be, corporeity, animal life, rationality, and a certain external form, which for distinction we call the human. Every existing thing which possessed all these attributes, would be called a man; and any thing which possessed none of them, or only one, or two, or even three of them without the fourth, would not be so called. (Mill 1843, 35.)

> On such an account, the meanings of our general terms are determined exclusively by the general characterizations. If we happen to talk about actual items in the world, it is because they happen to fit the general characterization. Of course Russell's theory of descriptions was eventually used to extend this connotative picture to proper names as well.

[21] One can occasionally find philosophers willing to bite this bullet as well. Katz, for instance, claims that "if Putnam's story about the things we have been calling cats turning out to be robots were true, it

that *none* of our applications of a given term are correct -- the connection between language and the world becomes more tenuous than it would intuitively seem. On the other hand, if the meaning of a term is determined solely by its putative extension,[22] then it becomes increasingly difficult to see how we could ever misapply it.[23]   Apparently substantive claims like "this is an X" must be understood as meaning roughly "this is a member of the set of things I call 'X.'" The latter being a much less interesting claim that can be made true simply by its sincere utterance. Compare the differences between "She is a witch" and "Now she is what I'd call a witch." The second, when sincerely uttered, might seem always to be true, but if meaning is determined solely by conditions of application, the truth-conditions of the first utterance must be understood in terms of the truth-conditions of the second.

Consider, for instance, the following two sentences:

(1) This is a cat.

(2) Cats are carnivorous.

If the meaning of "cat" is understood in terms of its general characterization, it is clear how a substantial claim can be made with (1), but a sincere utterance of (2) will merely be true in virtue of the meaning of "cat" (making it difficult to see how a substantial claim is made with it).[24] On the other hand, if the meaning of "cat" is understood in terms of its putative extension, then it is clear how a substantial claim can be made with (2), but sincere utterances of (1) acquire the same 'analytic' character that utterances of (2) did when

---

would follow simply that there never were any cats! There couldn't be, since to be a cat something must be an animal." (Katz 1975, 96.) (See also Lewis 1929, 26-7.)

[22] Indeed, Quine's notion of stimulus meaning does precisely this. Quine claims that the stimulus meaning of a sentence for a subject "sums up his dispositions to assent to or dissent from the sentence in response to present stimulation," and that stimulus synonymy "can be used as a standard of translation not only for observation sentences but for occasion sentences generally" (Quine 1960, 34, 217). While Davidson differs from Quine in focusing on the distal rather than proximal stimuli, he shares with Quine a tendency to identify what our words *refer* to with whatever we are disposed to apply those words to.

[23] This has become very apparent in discussion of the "problem of error" (or "the disjunction problem" (Fodor), "the misrepresentation problem" (Dretske) or the "normativity problem" (Millikan)) in the works of 'information semanticists' who take truth-conditions solely to reflect the putative extensions of the terms in question.

[24] Of course this conclusion can be moderated a bit with 'cluster' description theories such as Searle's, though the majority are still viewed as 'collectively' analytic. (See section 3.21.)

meaning was understood in terms of its general characterization.[25]　General characterizations and putative extensions are each good at explaining how the utterances that characterize the other are contentful, so any account that focuses on just one of the two will not be able to account for the contentfulness of the very utterances it takes to be constitutive of meaning.　The favored class of statements thus have the status conventionalists had assigned to the axioms of geometry.　Viewed from outside the framework, they collectively defined the framework itself, but from within the framework, they express claims that are necessarily true in virtue of the constituted meanings. There is a sense in which tying meaning to general characterizations allows one to explain how we manage to say something with our utterances, but prevents one from adequately explaining how what we say is about anything, while tying meaning to putative extensions allows one to explain how our utterances are about objects in the world, but prevents one from explaining how we manage to say anything about them.

Of course, few people have suggested that general characterizations and putative extensions determine the meaning of *all* our terms.　There is, however, a popular strategy of splitting the terms in the language into two groups, the first consisting of terms whose meaning is determined by their putative extension, the second consisting of terms whose meaning is determined by their general characterization.　Bertrand Russell could be understood as presenting a theory with roughly this form.　Logically proper names have their meaning determined by their putative extensions while descriptions, disguised or not, have their meaning determined by the associated general characterizations.[26]　This leaves

---

[25] Some have tried to make this failing into a virtue and parlayed an account's inability to account for objectivity in this area into a positive anti-skeptical virtue of the account. (For a discussion of this, see appendix B.)

[26] Indeed, something like this structure is found in many if not most theories with a sharp distinction between 'observation' and 'theoretical' terms. This two-tiered structure can also be attached to various causal theories, with names, natural kind terms etc. having their meanings determined by the putative extensions associated with a term's baptism and with the other terms having their meaning determined by general characterizations the components of which eventually reduce to terms of the information-bearing variety.

> We think that the description theory of reference fixing may well apply to 'pediatrician'; that it may well apply to non-basic "artefactual kind" terms; and that it may well apply to many other terms; for example, 'bachelor,' 'vixen' and 'hunter.' This is no threat to our theory of language, because the

the various types of 'analyticity' in place for different parts of the vocabulary, so while statements of type (1) could be made substantively with some terms, and statements of type (2) made substantively with others, there would be no terms for which statements of *both* type (1) and (2) could be made substantively. However, since "cat" seems to be used substantively in both (1) and (2), two-tiered theories do not capture our intuitions in this area.

A different type of 'hybrid' position equates a term's meaning with a *combination* of its putative extension and general characterization (the first picking out the 'wide' and the second the 'narrow' content). However, it will be argued that simply *equating* wide content with putative extensions would seem to get the truth-conditions of our terms wrong (see 3.22), but if narrow content is determined by the general characterizations, it would seem that wide content would have to be determined by the putative extensions.

3.2 Attempts to salvage objectivity from within a single component

While there are obvious problems when one simply *equates* a term's meaning with either its putative extension or its general characterization, it may seem that one could treat the term's meaning as *determined* by one of the two and still capture our intuitive notion of objectivity. By having meaning determined by, rather than equated with, the chosen meaning-constitutive aspect, one leaves room for the idea that elements within the meaning-constitutive aspect could still count as being mistaken.

3.21 General characterizations and inferential-role semantics

That accounts based upon general characterizations had problems accounting for objectivity was not lost on everyone within this tradition, and some (e.g. Block) simply gave up on the idea that the type of 'meaning' that was determined by our general

---

reference of all the terms appealed to in a description theory must ultimately be explained causally. (Devitt & Sterelny 1987, 78.)

Of course, it isn't at all clear that we really could find descriptions involving only natural kind terms which would correspond to every term in our language, and the suggestion that typical users of the language have the reference of their non-natural kind terms fixed this way seems incredible.

characterizations should, or could, concern itself with issues relating to truth.   A less pessimistic line would be to try to fund some notion of objectivity from *within* the general characterizations. This is the line taken by Searle when he moves from 'straight' to 'cluster' description theories in passages such as the following:

> Suppose we ask the users of the name 'Aristotle' to state what they regard as certain essential and established facts about him. Their answers would be a set of uniquely referring descriptive statements. Now what I am arguing is that the descriptive force of 'This is Aristotle' is to assert that a sufficient but so far unspecified number of these statements are true of this object.… suppose we were to drop 'Aristotle' and use, say, 'the teacher of Alexander,' then it is a necessary truth that the man referred to is Alexander's teacher -- but it is a contingent fact that Aristotle ever went into pedagogy (though I am suggesting it is a necessary fact that Aristotle has the logical sum, inclusive disjunction, of properties commonly attributed to him: any individual not having at least some of these properties could not be Aristotle).[27]

What a term refers to can be understood in terms of what most of the items in the general characterization are true of, so any single element in the general characterization can be false, even if their disjunction is analytically true.

Of course, such attempts to fund a notion of objectivity from within the general characterization have serious limitations, and Kripke has argued that, for instance, *all* of the general characterizations associated with a name could turn out to be false of its referent.[28]

3.22 Putative extensions and information-based semantics

Dissatisfaction with accounts of meaning and reference that focus excessively on the general characterizations became more and more evident from the early seventies on, since it became generally accepted that our specifications could often underdetermine or even misidentify what we intuitively took ourselves to be talking about. If anything, accounts of meaning and reference now pay excessive attention to the putative extensions.

However, as discussed above, simply *identifying* what a term refers to with its putative extension has the serious drawback of being unable to fund any substantial notion of objectivity with respect to obviously contingent utterances of ours such as "That is a cat." How then is any normativity to be salvaged from the dispositions that determine a term's

---

[27] Searle 1958, 94-5.
[28] See, for instance, his discussion of "Aristotle" in Kripke 1972.

putative extension? One popular strategy is to find some privileged set of dispositions that can be taken to be determinative of meaning, leaving open the possibility that the dispositions to apply our terms that are outside of the set will turn out to be misapplications. The two most popular candidates are those dispositions that accord with the dispositions of the community and those dispositions that we would have under certain 'optimal' conditions.[29] Neither of these, however, seems to adequately capture the normativity in question.

Consider first the 'communitarian' accounts that take the relevant dispositions to be those that agree with the dispositions of the community.[30] Such accounts require that, at the community level, a term's putative extension just *is* its actual extension, and such a requirement seems quite unintuitive. Entire communities can be, indeed often have been, mistaken about the proper extension of particular terms in their language, so the communitarian accounts cannot fund the sort of full-blooded conception of objectivity we want. Furthermore, even if one were willing to accept that the community's usage was incorrigible in this way (which is, of course, is an admission most of us are unwilling to make), the communitarian accounts only provide an explanation of an individual's *non-perceptual* errors. Bert can be understood as mistaken in claiming that he has arthritis in his thigh because his disposition to apply "arthritis" to pains in his thigh does not match the dispositions of the community. Perceptual errors are less well accounted for. If Bert is disposed to apply the term "deer" to cows seen at a distance with antlers strapped to their heads, there is no reason to think that the communitarians will be able to account for this sort of systematic perceptual error. If cows with horns strapped to their heads look a lot like deer from a distance, then it will be true that other members of Bert's community will be disposed to 'misapply' the term in the same way that he does.[31] (Of course, the

---

[29] Boghossian 1989, 534.
[30] Such a position has been attributed to both Burge 1979 and Kripke 1982, though both attributions are questionable.
[31] Boghossian 1989, 536.

community could tell that there was an error if it were allowed to 'take back' previous assertions when closer, but this type of story could be told about Bert on his own.)

The other alternative is to take the privileged set of dispositions to be those we would have under certain 'optimal' conditions (the conditions under which we learned the term, the conditions under which our cognitive mechanisms are functioning as they should, etc.).[32]   Such accounts have frequently focused upon accounting for putative 'misrepresentations' such as a frog's 'mistaking' a BB for a fly,[33] but it seems unlikely that any such account will be able to explain errors such as Linnaeus's applying the term "fish" to whales. In particular, the putative extension of "fish" for (pre-1776) Linnaeus may be no different from the putative extension of "fish" for Ishmael. The fact that one correctly applies the term to whales while the other is making a mistake can, it seems, only be explained in terms of differences between their general characterizations. As a result, attempts to account for normativity by supplementing putative extensions exclusively with extra-conceptual factors will be unable to account for the difference between the two. Furthermore, the extensive literature on this sort of proposal shows a fairly clear consensus that the proposals on the table so far are all inadequate for solving the 'disjunction problem.' That is to say, they all fail to provide an explanation of how a dispositional account of meaning can leave room for the person with the dispositions ever making a mistake.[34]

---

[32] The cognitive system's "functioning as it should" is here usually given some sort of 'teleological' explanation (ultimately to be explained by evolutionary biology). It should be noted, however, that one could understand a system as "functioning as it should" when all of its terms are at an equilibrium, though such a proposal would, of course, make essential use of the general characterizations.

[33] The frog's much discussed response to small black objects in its environment may be a paradigmatically punctate disposition, and if human misrepresentation is to be understood on analogy with such cases, then general characterizations will play no essential role in accounting for it.

[34] It might be thought that such a strategy could be supplemented by some type of 'metaphysical essentialism' to get at a more full-blown notion of objectivity. An agent's dispositions will pick out a certain putative extension, and the world's 'essential structure' will determine which items are part of the term's extension (i.e., those items that have the same essential structure as most of the privileged sample). Such accounts are discussed in greater detail in appendix A.

The dispositions that determine the putative extensions of our terms, like those that determine the general characterizations, cannot fund a sufficiently substantial notion of objectivity on their own, and so there is reason to look for another type of solution.

3.3 The need for a dynamic conception

Since we would like to see ourselves as making *substantive* claims *about* the world around us, a picture of meaning as the product of the *dynamic* between putative extensions and general characterizations seems desirable. *Saying* something *about* the world requires us to have both the putative extensions and the general characterizations in play for each word. Having each put checks on the other allows us to have a fairly robust conception of responsibility both to what it is for something to be an X (since aspects of a term's general characterization may be out of line with the majority of the items to which the term is applied), and to what items are, in fact, X's (since some of the items that the term is applied to may be out of line with deeply entrenched aspects of its general characterization). Any account of meaning in which a term's extension is determined by just one of the pair in question will be one in which there are no checks on the criteria in question, and will thus provide a comparatively anemic account of semantic normativity.

There may be other types of normativity that could be appealed to. Someone whose general characterization of a given word was satisfied by nothing could be viewed as violating the more 'pragmatic' norm that we should use words that actually pick out items in the world. In much the same way, someone whose putative extension for "gold" included iron pyrite could be seen as violating the pragmatic norm of having one's terms map on to natural kinds. In both cases the norms are pragmatic norms suggesting that we have made a 'mistake' by associating a particular meaning with a particular term, not semantic norms suggesting that we have made a mistake (a false belief or misapplication) *because* of what we mean by our term.

4. Equilibrium as a regulative ideal governing our linguistic practices

Our use of language is thus best understood as developing in a way that can be called, for want of a better word, "Socratic."[35] This practice is regulated by the idea that, over time, conflicts between a term's putative extension and gerneral characterization will be resolved, and that one will end up with an 'equilibrium' in which the general characterization matches those items within the putative extension.[36] In some sense, meaning itself could be understood as corresponding to what our linguistic use would be in such an 'ideal' state. The idea that we do, in fact, mean something by our terms regulates our practice of trying to eliminate the inconsistencies between our ascriptional practices and our general beliefs. Of course the move towards equilibrium could be driven in part simply by our desires to keep our beliefs consistent. If we believe that all X's are P, that a is an X, and that a is not P, then we will want to give up one of the three beliefs (the first of which could be part of our general characterization of X's, while the second could be part of "X"'s putative extension). Nevertheless, our response to these conflicts is structured not *just* by a desire for consistency. While consistency can always be produced by *either* modifying the general characterization or altering the putative extension, in most instances we see one of these two as what is required given what we mean by the term. Which we choose is determined, of course, by how well entrenched the incompatible factors are within our practice. Since we view the process as getting at what we 'really' mean, not any way of eliminating the inconsistency will do.

Our assumption that we mean anything at all involves something like the regulative use of an idea of reason. There may not necessarily be anything corresponding to it, but it guides our investigations towards an independently desirable equilibrium. We do not see

---

[35] One can see instances of this 'Socratic' method in Hegel's discussion of knowledge (Hegel 1807), Goodman's discussion of induction (Goodman 1954), Rawls' discussion of justice (Rawls 1971), and Burge's account of linguistic development in general (Burge 1986c).   Something like this methodology is suggested for semantic terms in Devitt 1994, 1996, though Devitt seems to build into his account a number of unjustified presuppositions about the methodology (Devitt's account and some problems with it are discussed in appendix C).

[36] It should be stressed that this ideal regulates the use of individual speakers as well.

ourselves as simply being disposed to apply words to things that may or may not fit out beliefs about them, rather we understand ourselves as speaking a (shared)[37] language in which there are matters of fact about which general beliefs are true and false and which items we apply our terms to genuinely fall within their extension.

4.1 Conflicting pictures of linguistic development

Since we are committed to making our general characterizations and putative extensions cohere together whenever we find that they are out of step with one another, it can be tempting to think that there is a rationally coherent structure behind our practice, a pre-existent and independent structure against which our practice is evaluated. The regulative idea of meaning thus encourages a comparatively simple picture of linguistic development. In this picture there are roughly three things involved in a term's development: what the term actually means (what it refers to, its *actual* extension), what we take the term to apply to (its *putative* extension), and what we believe about what the term stands for. The simple picture sees linguistic development as arising when we realize that the putative extension of a given word doesn't match its actual extension and adjust our usage accordingly. For instance, when I was younger, I would have fairly comfortably classified mushrooms as a type of vegetable. When asked what vegetables I didn't like, mushrooms would have been near the top of my list, etc. I used to call mushrooms "vegetables," but I was eventually informed that mushrooms weren't vegetables, so I stopped calling them "vegetables." It can be tempting to see this part of my linguistic development as my simply bringing the putative extension of "vegetable" into line with the actual extension. According to this picture, linguistic development is merely a matter of chipping away at various misapplications of our words until we discover their 'true' extensions. Much the same can be said of the beliefs: we have a number of present beliefs, and we alter them so that they

---

[37] That we understand this language as shared is fairly important in explaining how we coordinate our language use with that of others. The suggestion that the idea of a shared language functions as a regulative ideal controlling our linguistic behavior is not only more phenomenologically plausible than more voluntaristic explanations such as Lewis's account of convention, but is also compatible with non-individualistic theories of content in a way in which conventionalist accounts are not. (For a more extended discussion of these issues see my "Convention and Language.")

correspond to what is actually true of the referent of the term in question. Such a picture is perhaps most compelling with natural kind terms, but it lies behind our thought about other tracts of our language as well.

We run into problems if we come to assume that there is something 'out there' corresponding to this regulative ideal, and the natural thought that both the conditions of application and the general characterizations are subject to correction can lead to a quasi-platonistic conception of language in which linguistic norms are robustly independent from our actual linguistic practice. (Indeed, something like this platonizing picture can be seen as motivating many, but by no means all, complaints to the effect that our language use is 'degenerating' or is being 'corrupted.') This quasi-platonist idea that meanings inhabit some practice-independent realm encourages the idea that they are available to us through some special 'intuition' and can thus be determined through introspection or some sort of 'conceptual analysis.'

Such hopes are destined to be disappointed, and the fact that we cannot find anything in experience that corresponds to this regulative ideal can encourage another picture of linguistic development that goes to the opposite extreme. The rejection of the quasi-platonistic picture of linguistic norms, in which meanings are 'out there' and our practice is responsible to them, can leave one with a type of skepticism about the possibility of any sort of linguistic normativity that criticizes the practice itself. (I.e. it assumes that to criticize our practice we must be going beyond the practice itself.)[38] What we mean should be determined by our linguistic practice, so it may seem, at times, hard to understand what could constitute the 'actual' extensions that stand over-and-against the putative ones (which seem to reflect our practice), and if the independently determined extensions can't be made sense of, they drop out of the picture, leaving the putative extensions on their own. What a speaker means by a term is understood, on such accounts, in terms of how he is disposed

---

[38] One way to correct the practice of the individual is to appeal to the practice of the surrounding community, but not only does this leave the objectivity problem at the communal level, but such accounts can seem like the arbitrary imposition of social usage upon the individual.

to apply it. Any attempt to get a more disposition-transcendent notion of meaning can thus come to be viewed as a type of unmotivated supernaturalism.[39]

Such an account simply denies that linguistic development occurs at all. There may be linguistic *change*, but that is all. For instance, when I stopped calling mushrooms "vegetables," I just changed what I meant by "vegetable." While this picture has been defended even for natural kind terms, it is at its most plausible for non-natural kind terms. For instance, a change in usage seems more naturally explained by our coming to mean something different by, say, "chair," than it is by our discovery that some things we have always called chairs weren't *really* chairs at all. Meaning is here equated with an additive rather than a dynamic combination of the general characterizations and putative extensions, and any change in either will amount to a change in meaning. This view gets at least some of its plausibility from the problems with the more platonistic picture, but it has serious problems of its own.[40]

Not only does it seem like an unintuitve story for natural kind terms, but it is also by no means obviously true for terms like "chair" either. For instance, a speaker can, being fully aware of his own past usage and the linguistic usage in his community, still argue that some of the items that the community calls "chairs" are, in fact, stools. If he provides good reasons for his claim (the pieces of furniture in question are structurally much closer to prototypical stools than to prototypical chairs, the furniture in question was originally made and marketed by a speaker who was notoriously bad at distinguishing stools from chairs, etc.), then it is not an acceptable reply to his criticism to say, "Listen, they must be chairs, because that's what everyone in the community calls them." Indeed, I think that we see something very much like this when certain high-backed bar-stools are assumed to be

---

[39] Davidson, following Bilgrami, claims that "the yearning for such norms is a nostalgic hangover from dependence on a Platonic conception of meaning." (Davidson 1993a, 145.) See also: "Unless meanings are laid up in heaven, they must depend on the actual practice of someone or some group." (Davidson 1993b, 117.)

[40] Still, it should be clear that having all of this information will allow us to 'communicate' with the interpretee in the sense of being able to get him to do things, to retrieve information about the world from him, etc.

stools (a natural inference given what they are called), though they are probably more accurately characterized as chairs. Such accounts can't understand content in terms of truth, but rather, just truth as the agent conceives it,[41] and it isn't clear that one can take this latter notion as primitive. After all taking something to be true seems to presuppose the possibility of one's being mistaken, but if what one means is equated with what one takes to be true, such a possibility is ruled out.

However, such pessimism about linguistic norms is unwarranted. Both views discussed above assume presuppose that a practice must be criticized by something outside of it if it is to be criticized at all. However, the 'logical' requirement that one's beliefs and applications be consistent can provide 'internal' grounds for criticizing a practice, and thus will make it possible for one to make non-perceptual errors with one's terms. One can criticize linguistic practice without going 'beyond' it.

One thus need not look beyond our own linguistic practice to insist that what we mean by a term corresponds to how we *should* be (rather than *are*) disposed to apply it. The normativity of the "should" comes from the need to resolve any tensions between *our own* dispositions and beliefs. There is thus nothing mysterious about the norms involved. Lexical norms come merely from consistency constraints upon our attitudes towards a term's general characterization and putative extensions. Lexical norms follow from the norms pertaining to consistency that come from understanding an agent's utterances in terms of truth. Present usage can be criticized by reference to its divergence from an equilibrium that is itself determined by (but not equated with) present usage.

---

[41] This is fairly explicit in Bilgrami's account, which claims to make no use of notions like truth or reference.

4.2 On sense and meaning

At this point it might be worth discussing how the picture outlined above maps on to the traditional sense/meaning distinction.[42] (What follows should not, however, be taken as a piece of Frege exegesis). On the account above, we can correlate a term's sense (roughly, its *cognitive* value) with its present putative extension and general characterization, while we can correlate its meaning (its semantic value) with the equilibrium determined by its sense. This allows the sense to be, in a fairly intuitive sense, the mode in which the meaning is presented. It has the truth conditions of a speaker's utterances to determined by the meanings of his terms, while still allowing the truth of *de dicto* belief ascriptions to be determined by the senses of the terms involved.[43] Since equilibria are a function of the terms putative extension and general characterization, there remains a clear sense in which this picture allows that sense determines meaning. Once again, this is not meant to be a piece of Frege exegesis, but is only meant to illustrate how the framework outlined above can capture much of the intuitive distinction that so many have found appealing in Frege's essay.

5. How there might not be any (or any *single*) endpoint

That there might not be anything corresponding to the regulative ideal becomes clear when we consider cases where there seem to be many or no equilibria available. Given a set of putative extensions and general characterizations there will, of course, always be many potential equilibria, since if there is a conflict between the putative extension and the general characterizations, it will always be possible to remove the inconsistency by altering either of the two sides. However, an equilibrium is "accessible" if it is does not require giving up elements that are too deeply entrenched, and not every *potential* equilibrium will be an *accessible* one. It should also be noted that how much can be given up will also

---

[42] See Frege 1892.
[43] In a sense that will be made clear in chapter three, the senses of a speaker's terms will determine what substitutions he will be willing to accept, and thus which *de dicto* attributions will be true of him.

depend, to a certain extent, upon how much must be given up to reach alternative equilibria. If there is a possible equilibrium that requires giving up next to nothing, it can make what would otherwise be accessible equilibria inaccessible. The comparative entrenchment of various elements in the pre-equilibrium state will make various potential equilibria inaccessible. For instance, until quite recently I believed that fish never give birth to live young. It was pointed out to me that guppies (which I took to be fish) do give birth to live young. I could then have made my beliefs consistent in two ways, give up the belief that fish never gave birth to live young or stop calling guppies fish. I would have taken the second of these options to be perverse. Not because there would be something 'objectively' wrong about making such a decision, but rather because (like most, if not all, people) I do not have the "don't give birth to live young" belief nearly as deeply entrenched as the belief that guppies are fish. More often than not, we will feel that one particular way of resolving the inconsistency is the one that gets at what 'we really meant all along.' Nevertheless, there are occasionally instances where a number of possible resolutions seem equally accessible (the debate over whether "fish" should have been treated as a natural or functional kind term may have been one of these cases), and there are other cases where there seem to be no acceptable equilibria and either an 'unperceived ambiguity' may have to be posited (Field's discussion of "mass" might be an instance of this), or the term may be given up as unusable.

6. How to understand the dialectical 'development'

The dialectic outlined above can be understood as describing how we come, over time, to discover what our terms have always meant, as describing how what we mean by our terms comes to change over time, or even as how, over time, we come to mean anything (determinate) at all (with nothing (determinate) actually being meant until the equilibrium is

reached).[44] For terms not already at an equilibrium (which may be, for all we know, most of our terms) we are left with the question of whether what we (now) mean by them should be identified with the present stage in the dialectic, or where we hope the dialectic to be heading. Which of the two alternative descriptions we should use has been the subject of considerable, and often inconclusive debate.[45]

It is difficult to decide whether such a process should be described as our changing what we mean or as our changing our conception of what we always meant, because "meaning"[46] is precisely one of those terms for which we have not even come close to reaching any sort of equilibrium. Not only are many of our general intuitions about language and meaning in conflict with each other, but these ideas are often in conflict with the way we actually apply semantic terminology.[47] The presence of conflicting intuitions and practices associated with our semantic terms gives proponents of the alternative ways of describing linguistic change material to support their claims about whether or not the meaning of our terms has changed through the dialectical development of our linguistic practice.

As a result, a good deal of the contemporary discussion about thought and linguistic content can be usefully understood along the general model of linguistic development outlined above.   That is to say, there is, roughly, a set of general beliefs we have relating to meaning and content, and our semantic terms have a putative extension. These do not involve cases of saying "look at that meaning" or "did you see that content over there?" in

---

[44] I know of no one who defends the extreme version of this view, though Field's account of 'partial reference' could be understood as a version of the more moderate alternative which understands the dialectic as describing how what we mean becomes more *determinate* over time. (See Field 1973, 1974.)

[45] This debate has been, of course, most prominent with respect to theoretical terms in science (with writers like Feyerabend (1975) advocating the former type of description and others like Putnam (circa 1975) advocating the latter), but the same issues carry over to other types of terms as well.

[46] As is the case with "reference," "belief," and other such semantically loaded terms.

[47] See Kay 1983 for the claim that like many other 'folk' theories, our tacit theory of language is not globally consistent and incorporates conflicting conceptions of what language is.

the way that the putative extension of "cat" would depend upon our utterances of "that's a cat," "look at that cat" or "did you see that cat over there?" Rather they involve our 'standard' or 'commonsense' ascriptions of content, and similar evaluations of the meanings or truth-conditions of the statements of others.[48] Much of the most interesting work in the philosophy of mind and language has arisen from the gradual realization of just how much tension there is between our general beliefs about meanings and contents and our standard practices of ascribing them.

If the general picture of linguistic development outlined above is right, and it also applies to our theorizing about ascriptions of meaning and content, then one should not expect knock-down arguments that will leave everyone, if not happy, at least convinced. Thinking that the debate will necessarily have this character is the result of thinking that there is some antecedent truth about meaning and content and that we just have to sniff around enough until we find it (i.e., some form of the platonist picture.) Resolving the tensions between our ascriptional practices and theoretical commitments may involve some give and take, and it thus may involve giving up things that would be, all things being equal, worth having. (Again, the parallel with ethical/practical deliberation is worth noting.) Because of this, it is important not to motivate a particular view by something like the argument from elimination (i.e.: by pointing out that all the other positions force us to give up something which we should reasonably want), since it seems quite probable that there is no position that would not force us to give up something which we would reasonably want.

Furthermore, when evaluating these conflicts between the ascriptional practices and general beliefs, *all* the general beliefs should be taken into consideration, especially since the tension between the ascriptional practices and the general belief may be the result of

---

[48] For a similar appeal to content ascriptions to fill such a role, see Devitt 1994, 1996. Of course, this notion of just what the 'natural' ascriptions of content are (or even the notion that there is such a thing) is not without its problems.

tensions among the general beliefs themselves.[49] Indeed, this process of tension resolution allows us to look at things on, as it were, a more global level. That is to say, when resolving a tension in one area, general beliefs we have in other areas can be taken into consideration. This looking towards a 'total package' of views can considerably alter how things work out at a local level.

In the philosophy of language, which often considers itself the 'foundational' part of philosophy,[50] one can occasionally get the impression that it is 'not fair' to bring up considerations from other areas of philosophy when evaluating a semantic theory. For instance, while a theory of meaning might be rejected if it entailed that we could never learn or know a language, if it entailed that we could never know anything about, say, the past, or even the world around us, such a result might instead just be considered an interesting (if disturbing) consequence.[51] This attitude isn't really defensible if the general view presented above about what is involved in settling issues about meaning is right. If one's account of content entails that we can't know anything about things we intuitively take ourselves to know, then this counts as an at least *prima facie* reason to think that one's account is wrong.

7. Some general beliefs about meaning

There are a number of general beliefs we have about meaning that non-philosophers rarely attempt to formulate explicitly, and not much should be taken to hang upon the

---

[49] The fact that the general beliefs that "fish" picked out a natural and a functional kind were inconsistent was made manifest by the discovery that whales, while of the same 'functional' kind as other fish, were not of the same 'natural' one.

[50] "The theory of meaning is the fundamental part of philosophy which underlies all others." (Dummett 1973, 669.)

[51] The fact that the classical theory of ideas was held so long in spite of its clearly disastrous consequences for epistemology might be an instance of this. Michael Dummett's work could also be understood this way, and the sentiment is aptly expressed, though not endorsed, by Fodor and LePore in the following:

> Serious though these worries are, however, they are all, as it were, *external* to functional role semantics as such; they impugn its consequences for epistemology and ontology rather than its coherence. And a convinced New Testament Semanticist might be prepared to bite the bullet. If the doctrine that meaning supervenes on intra-linguistic relations has relativistic, idealistic or even solipsistic consequences, then perhaps we had better learn to live with relativism, idealism or even solipsism. (Fodor and LePore 1993, 20.)

particular formulation of them here. For instance, it seems clear that we take our words to refer to (and our thoughts to be about), independent objects which they represent to a greater or lesser degree of accuracy. This 'belief'[52] is fairly deeply entrenched, and will be formulated for present purposes as:

(1) Our beliefs and sayings are true or false of (about) the world around us.

There are a number of 'philosophical' beliefs about meaning and content which while perhaps not widely shared, owe their plausibility to their status as either explications of (1), or things that must be true if (1) is to be.[53]

Another deeply entrenched belief has to do with what is often referred to as "self-knowledge," "first-person authority," "privileged access" etc. We tend to know better than others do, and seem to know in a different ("unmediated") way, what we believe. That is to say:

(2) We tend to know what we are thinking.[54]

A further belief that is equally entrenched is that our behavior can be explained by our beliefs and desires. Once again, while this belief is often not explicitly formulated, it is clear that we take an agent's beliefs to be relevant to his behavior. For the time being the belief can be expressed as:

(3) Our behavior is explained by our beliefs and desires.[55]

---

[52] While formulated here as explicit beliefs, some of these 'beliefs' might better be thought of as part of a, possibly pre-intentional, "Background" of abilities and disposition that can have a lot to do with what type of equilibrium can be reached for these terms. (For a fuller discussion of this, see appendix F.)

[53] For instance, the belief that we refer to objects in virtue of being able to specify (through demonstratives, descriptions etc.) what those objects are can be understood as an (unsuccessful) attempt to explain how we refer to items in the world. The belief that our terms have determinate extensions is related to (1) and may even be a relatively philosophical explication of it.[54] Like (1), there is no shortage of philosophical beliefs trying to make sense of (2). [55] This should not, of course, be taken as an empirical generalization about what we actually say when we explain someone's behavior. There is, of course, considerable debate about how (3) should be explicated, in particular, about what type of explanation is involved (causal, lawlike, hermeneutic, etc.).

Other beliefs that could be seen as part of the general characterizations of our semantic terms are the following:

(4) When an agent sincerely utters a sentence with a certain content, he has a belief with the content of his utterance.[56]

(5) There is a fact of the matter about what we mean.[57]

(6) We speak a shared language.[58]

While it is undoubtedly the case that these more philosophical characterizations of semantic content are more easily given up when they conflict with our ascriptional practices, it should not be assumed that this is always so. Since the philosophical characterizations are often the product of reflection upon (and taken to be an explication of) the more primitive beliefs, giving up some of the philosophical beliefs may require giving up some of the primitive ones (or at least coming up with a new understanding of them).

For instance, methodological solipsism[59] (roughly the idea that the meanings of our words and the contents of our thoughts could be characterized independently of our relations to items and events outside of our heads) may seem like a comparatively recent philosophical prejudice about meaning that we have no reason to keep it if it conflicts with our ascriptional practices. It provides, however, a fairly elegant way of accounting for many of the more primitive beliefs.[60] It seems initially to do well with behavioral explanation, the factuality of meaning, self-knowledge, and to provide an account both of how language relates to the world, and of how we are related to language. When methodological solipsism was given up, many of these more primitive beliefs were questioned (we don't really know what we mean, etc.), and, indeed, it became quite popular to suggest that there is a methodologically solipsistic 'narrow' content that is, in

---

[56] This belief and various attempts to explicate it are the subject of chapter two.
[57] Given how one feels about facts, this can lead to the more obviously philosophical belief, "Semantics is an empirical science like any other" (Devitt 1994, 545).
[58] Which has associated with it more philosophical beliefs such as that meanings are necessarily public, and that meaning depends in part upon convention.
[59] A position defended most vigorously in Fodor 1981, 1987. [60] Though not all. In particular, the idea that we speak a shared language drops out.

some sense, independent of our ascriptional practices and used to make true the general beliefs.[61]

8. Attitude ascriptions

It was suggested above that attitude ascriptions can be treated as evidence for the conditions of application of our semantic terms. The role of 'thought experiments' of the sort made famous by Kripke, Putnam, and Burge can be understood as a way of eliciting such data about how semantic terms are actually applied. Understood this way, their thought experiments are comparatively unambitious: they merely draw our attention to the type of ascriptions we are inclined to make, but they do not commit us to the truth of such ascriptions (they relate to what we *do* say in such situations rather than what we *should* say). Of course the authors believe that what we do say is, in fact, what we should say in these cases, but even those who do not share their intuitions about the thought experiments can agree that most people will attribute content the way the authors suggest.

It should also be stressed that, when saying that the ascriptions themselves are fairly uncontroversial, one is saying that people faced with either of the two possibilities that, say, Burge describes will ascribe content in the way Burge says they would. It does not involve saying that most people, when faced with a thought experiment involving *both* possibilities, will describe the pair the way Burge does. Being faced with such contrasting pairs may cause us to bring our general beliefs into play, and in such cases, our ascriptions are more likely to reflect our general beliefs about content than our general ascriptional practices. That is to say, it does seem true that people when faced with the

---

[61] Some more philosophical beliefs relating to our semantic terms that will come up in the course of the dissertation are:

  i. Russell's principle: "a subject cannot make a judgment about something unless he knows which object his judgment is about." (Evans 1982, 89.)

  ii: The Unity Constraint: The contents that are evaluated as true or false are the very same contents that explain behavior. (Bilgrami 1992.)

  iii: If the speaker intends to use a name in the same way as it was transmitted to him then the object referred to by the speaker with the term is the same as the object referred to by the transmitter with the term. (Kripke 1972.)

arthritis/tharthritis cases separately will ascribe different contents to Bert and his counterpart, but it need not follow that people who are faced with both Bert and his counterpart will say that they mean anything different by "arthritis."

Four writers who have pointed out how certain aspects of our ascriptional practice were in tension with general beliefs we had about meaning are Saul Kripke, Hilary Putnam, Tyler Burge and Mark Wilson. They do not necessarily present themselves as doing this, but understanding them in this way does, I believe, no violence to their work. Still, this is a revisionary history of what is going on, not an exposition of what the four authors took themselves to be doing. While these writers all present us with examples of how our ascriptional practices can conflict with general beliefs about meaning, the ascriptional practices become increasingly easy to give up as one moves from Kripke to Wilson (or, correspondingly, the general beliefs become progressively harder to give up). Indeed, it is not surprising that Burge's and Putnam's thought experiments are the ones to have generated the most controversy. In the Kripke cases, we stick too closely to our ascriptional practices, in the Wilson cases, too closely to our general beliefs for controversy easily to get off the ground.

As stressed in section six, in evaluating how these tensions are to be resolved, it is important to get a complete picture of what is involved. Resolving tensions in a certain way locally may turn out to be incompatible with the most attractive global resolution. This involves first getting a thorough account of what the ascriptional practices are, since they often get much of their force collectively. For instance, while the ascriptional practices described by Burge may initially seem to conflict too much with our general beliefs to be accepted, once these general beliefs are modified to account for the Kripke and Putnam cases, the degree of conflict may be considerably reduced. I doubt that the intuitions behind Burge's thought experiments would carry the force they do if the general beliefs they cut against hadn't already been shaken by Putnam's and Kripke's work. Indeed, I would guess that before 1972, Burge's thought experiment would have been something of

a non-starter. It would have been too natural for philosophers to say Bert just meant by "arthritis" "a disease of the joints and thigh." It was only because such descriptive theories had been discredited already that such a response came to be a much less appealing option. In much the same way, sticking with the ascriptional practices in the Wilson cases would not seem even a remote possibility if one were not familiar with the Burge cases, which are, in many respects, structurally similar.

Each of the four authors highlights aspects of our ascriptional practice that suggest that some set of criteria that might have been assumed adequate for determining reference or content, can, in fact, underdetermine and even misidentify what we intuitively take ourselves to be referring to. In order to accommodate the relevant attributions, other factors -- such as the speakers' linguistic history, physical environment, social environment and even future history of use -- must be incorporated into the proposed set of criteria. The following will highlight the types of underdetermination and misidentification these authors bring to our attention.

8.1 Kripke: proper names and history of use

The first, and in some sense paradigm case of such a conflict is Kripke's discussion of proper names. The bit of our ascriptional practice he pointed to was our habit of interpreting people who use a proper name as referring to the historical source of most of the information associated with that name.[62] The general belief that Kripke explicitly targeted was the idea that the reference of a speaker's use of a name was determined by the descriptions he associated with that name, though little would have changed if other sorts of recognitional capacities were added as well. The types of problems Kripke raises against the descriptive theory of names (that the 'internal' information can often underdetermine or even misidentify the referent) will turn out to apply to internalist accounts of many other terms as well.

---

[62] This particular way of formulating just what Kripke pointed to owes more to Evans' discussion of Kripke (in Evans 1973) than to Kripke's own discussion.

Underdetermination: Feynman

The first of these problems relates to the fact that, while it may be the case that, for some names, we can come up with a set of descriptions which, in fact, is satisfied uniquely by the intended referent, the descriptions we associate with a name often radically underdetermine who that name refers to. Kripke gives an extended example of this sort of underdetermination in his discussion of our term "Cicero," but the general point is summed up quite nicely in the following discussion of "Feynman."

> Consider Richard Feynman, to whom many of us are able to refer. He is a leading contemporary theoretical physicist. Everyone *here* (I'm sure!) can state the contents of one of Feynman's theories so as to differentiate him from Gell-Mann. However, the man in the street, not possessing these abilities, may still use the name 'Feynman.' When asked he will say: well he's a physicist or something. He may not think that this picks out anyone uniquely. I still think he uses the name 'Feynman' as a name for Feynman.[63]

The basic point here should be clear. Kripke draws our attention to the fact that we are typically taken to be speaking about people we have no way, if left to our own resources (i.e., we cannot help ourselves to expertise of other speakers, written information, etc.), of uniquely identifying (other than using the name itself).

Misidentification: Einstein, Thales

The second problem is even more serious than the first. Not only is it the case that the descriptions a speaker associates with a name may fail to pick out its referent uniquely, these descriptions may actually pick out someone other than the person who the term is intuitively taken to refer to. This can be seen in Kripke's more extended discussion of names like "Gödel," "Aristotle," but is summed up quite clearly in the following paragraph:

> Even worse misconceptions, of course, occur to the layman. In a previous example I supposed people to identify Einstein by reference to his work on relativity. Actually, I often used to hear that Einstein's most famous achievement was the invention of the atomic bomb. So when we refer to Einstein, we refer to the inventor of the atomic bomb. But this is not so. Columbus was the first man to realize that the earth was round. He was also the first European to land in the western hemisphere. Probably none of these things are true, and therefore, when people use the term 'Columbus' they really refer to

---

[63] Kripke 1972, 81.

some Greek if they use the roundness of the earth, or to some Norseman, perhaps, if they use the 'discovery of America.' But they don't.[64]

Such cases convinced most philosophers that the descriptive theory of names was deeply flawed, and it seemed that other internalist accounts of proper name meaning would do no better.[65] What it was about us that seemed to relate us to the bearers of the names we used was not just the descriptions or the discriminating capacities we had, but also our being part of historical chains involving the use of each particular word.[66] Of course it is always open to the defender of the descriptive theory of names to be hard-headed (the no counterexamples, only consequences approach), and claim that we should just understand this aspect of our ascriptional practice as confused, but it is notable that, to my knowledge, no one seems to take this line.

8.2 Putnam: natural kinds and the physical environment

In much the same way, Putnam's Twin-Earth stories can be understood as showing that our ascriptional practices involving natural kind terms are in conflict with the idea that what we mean supervenes on anything like our brain states.

Underdetermination: beech, elm

Putnam's paper is full of examples of underdetermination, and one of the upshots of his Twin-Earth examples is that underdetermination is, at least in principle, possible even for familiar terms such as "water" or "gold," though he gives a more 'down to earth' example with his own use of "beech" and "elm":

> Suppose you are like me and cannot tell an elm from a beech tree. We still say that the extension of 'elm' in my idiolect is the same as the extension of 'elm' in anyone else's, viz., the set of all elm trees, and that the set of all beech trees is the extension of 'beech' in *both* of our idiolects. Thus 'elm' in my idiolect has a different extension from 'beech' in your idiolect (as it should).[67]

---

[64] Kripke 1972, 85. Much the same point is made by Donnellan, who, even if it were the case that the only description we associate with "Thales" is "the Greek philosopher who held that all is water." (Donnellan 1972, 374.)[65] Though, for a dissenting opinion, see Searle 1983.

[66] For another expression of this idea, see Geach 1972, 155.

[67] Putnam 1975, 226.

Once again, it seems that we treat people as referring to things in the world independently of their having any descriptions or discriminative capacity that would allow them to pick out the objects in question.

Misidentification: unicorns

Cases where the descriptions associated with a natural kind actually fail to be true of the kind in question but are true of a different kind do not appear in Putnam's paper, but, given what has been laid out so far, they are not hard to construct. Indeed, Kripke's discussion of "unicorn" provides such a case:

> The mere discovery of animals with the properties attributed to unicorns in the myth would by no means show that these were the animals that the myths were about: perhaps the myths were spun out of whole cloth, and the fact that animals with the same appearance actually existed was mere coincidence. In that case, we cannot say that the unicorns of myth really existed; we must also establish a historical connection that shows that the myth was *about* those animals.[68]

Further examples are easy enough to construct, and the general point remains clear. We tend to treat the physical kinds that we are historically connected with as affecting what we mean by our words in ways that cannot be accounted for within an descriptivist framework.

Putnam's examples are another instance where the general belief has tended to cave in to the ascriptional practice, but it is noticeable that in this case there has been considerably more resistance. Not only have there been many attempts to have it both ways (narrow content matches general beliefs, wide content matches ascriptional practice, etc.),[69] but there have also been those who argue that the ascriptional practice of saying that Caesar and his Twin-Earth counterpart meant something different by "water" is simply mistaken, and should be revised in light of our general beliefs.[70]

---

68 Kripke 1972, 157.
69 This reaction was, of course, made to Kripke's work too.
70 It should be noted that the Putnam and Kripke examples paved the way for a fairly enthusiastic reworking of the general beliefs about meaning in terms of things like "causation." One can thus see the emergence of a new framework that may conflict with other aspects of our ascriptional practice. For instance, our 'natural' ascriptional practices would be to ascribe contentful thoughts and sayings to Davidson's 'Swampman,' and while most of the individualist and internalist frameworks can make sense of this practice, the causal/historical framework cannot. As a result, Davidson argues that people who attribute thoughts to the Swampman are confused (see Davidson 1987), while others have taken

8.3 Burge: non-natural kinds and the social environment

Tyler Burge's much discussed example of Bert and his use of "arthritis" suggests that our ascriptional practice is in conflict with the idea that the contents of our thoughts and the meanings of our words are, in principle, independent of the contents of the thoughts and meanings of words of those around us.[71]

Misidentification: arthritis

While the sections on proper names and natural kinds began with underdetermination and moved on to misidentification, this one will do the opposite. This is partly because Burge's thought experiments usually involve cases of misidentification, and this should not be surprising. If one or two of the beliefs I associate with a given name turn out to be false, it is quite unlikely that there will be a person out there who does satisfy all the descriptions associated with the name. On the other hand, if some of my beliefs about, say, chairs happen to be false (for instance, I believe that they must have four legs and be made of wood), there will still be a recognizable class of objects about which all my 'chair-beliefs' will be true.

By far the most famous case of such misidentification involves Bert and his claim that he has "arthritis" in his thigh.[72] Bert has a number of correct beliefs about arthritis such as that he has had arthritis for years, that older people are more likely to have it, that stiffening joints is a symptom of arthritis, etc., but he also has the false belief that his arthritis has spread to his thigh. When he tells this to his doctor, the doctor tells him that arthritis is specifically an inflammation of joints, so Bert gives up his belief and goes on to wonder

---

the Swampman to show that there is something deeply wrong with the general beliefs about meaning that make up the causal/historical framework (see Fodor 1994).

[71] That is to say, while the contents of our thoughts etc. are *causally* dependent upon the contents of those around us (since it is our interaction with them that causes us to have the properties that constitute our contents), there is no *constitutive* relation between our contents and theirs.

[72] See Burge 1979, 77 and following.

what could be wrong with his thigh. In spite of his false belief, Bert is still taken to refer to *arthritis* by "arthritis."

Burge now asks us to consider a counterfactual situation in which Bert "proceeds from birth through the same course of physical events that he actually does, right to and including the time at which he first reports his fear to his doctor,"[73] the only difference being that, in the counterfactual situation, Bert's society uses the word "arthritis" to pick out inflammations outside of the joints as well, particularly in the thigh (this cluster of inflammations will be referred to as "tharthritis"). In such a case, Bert would be taken as expressing a true belief when he said "I have arthritis in my thigh," and that is because he taken to be referring to *tharthritis* by his word "arthritis."

These cases suggest that what we take Bert to mean by the word "arthritis" is determined not just by facts about his internal make up (or such facts plus facts about the purely causal relations between him and his environment), but also by the use of the term in the society to which he belongs.

Underdetermination: arthritis again

Burge does not really focus on cases that involve underdetermination, but given the scenario already in play, such a case is easy enough to construct.   What distinguishes tharthritis from arthritis is that one can get the former, but not the latter, in one's thigh, so a case of underdetermination can be seen if we consider a speaker much like Bert but with the qualification that he has absolutely no opinion about whether or not one can get "arthritis" in one's thigh. His beliefs would thus match equally well with either arthritis or tharthritis, but in each case one would attribute to him the concept used by the surrounding community (indeed, the impulse to do this is much stronger in the cases of underdetermination than it is in cases of misidentification).

---

73 Burge 1979, 77.

Much more so than with the Putnam and especially the Kripke cases, people[74] when faced with the practices Burge appeals to are willing to simply stick by the general belief and say that this part of our ascriptional practice is 'confused' or 'sloppy.'

8.4 Wilson: future usage

Finally, there is a range of cases brought to our attention by Mark Wilson that suggest that our everyday ascriptions of content not only reflect a sensitivity to the agent's internal make up, and the structure of his physical and social environment (including the past history of use of the terms he is using), but also reflect 'accidental' developments in the term's use in the time subsequent to any particular utterance.

Underdetermination: Grant's Zebra

For instance, around 1820 a fellow named Grant 'discovered' a purported species of Zebra native to Kenya and called it "Grant's Zebra." A few years later, the term "Chapman's Zebra" was introduced for a morphologically distinct type of Zebra found in what was then Rhodesia. Later still it was discovered that the two purported species of Zebra interbred near the Zambezi river and that, morphologically, one gradually faded into the other. Grant was then taken to have discovered a subspecies of the larger species *Equus burchilli*. This story is typical of the way we describe such cases. However, this aspect of our ascriptional practice is actually at odds with the general belief we have that what we mean by our words is independent of accidental features of our society's *future* linguistic practice. This general belief is in conflict with our practice in the Grant's Zebra case because, on reflection, it seems to be merely a *historical accident* that the term "Grant's Zebra" came to have the extension it did. If the taxonomists had investigated the area around the Zambezi river *before* they hit deepest Rhodesia, then they probably would have 'discovered' that Grant's Zebra could be found through most of East-Africa gradually changing into a different subspecies as it drifted towards Zimbabwe. In such a case, "Grant's Zebra" would have picked out the entire species, not just the subspecies found in

---

[74] For instance, Davidson, Bilgrami, Camp, my long past self.

Kenya. This is, of course, just a single example, but Wilson maintains that similar examples can be found "in virtually every case of enlargement of our world view through scientific progress."[75]

Misidentification: natural and functional kinds

It seems clear that many of Putnam's remarks about natural kinds could be understood as saying that a society's present usage can misidentify what a term refers to because there can be a diachronic as well as a synchronic division of linguistic labor. However, the fact that the division of labor can stretch out diachronically need not violate the belief that what a term means is determined by contemporaneous linguistic practice. Our present practice may make it clear that a term is a natural kind term and what constitutes being a natural kind, and if this is the case, how we would react to, say, certain chemical discoveries in the future may already be determined. The cases Wilson has in mind are more radical. In these cases it is the subsequent history itself that determines whether a term is a natural rather than a functional kind term, and so can also determine that large ranges of previous application (which would have been correct if the term had turned out to be a functional kind term) were incorrect. For instance, when the term "gold" (or its equivalent) was introduced, the introducers might have thought (to the extent that they thought about it at all) that the term picked out both a natural and a functional kind. With the discovery that the class of 'gold-looking' things contained more than one kind, one had the choice of giving up either the belief that 'gold' picked out a natural kind, or the belief that it picked out a functional one. The tradition of which we are a part chose to preserve the belief that it was a natural kind, and so early members of this tradition who called iron pyrite "gold" can be taken to be mistaken, even if they had no way of distinguishing the two. Had the tradition gone on to treat "gold" as a functional kind, the very same speaker would have

---

[75] Wilson 1982, 572. [76] This would, I believe, be the upshot of Field's claims about the 'Newtonian' use of the term (see, for instance, Field 1973).

42

said something true (though he might have had the false belief that "gold" picked out a naturally uniform kind).

Nevertheless, the conflict we see in the Wilson cases is one in which many, if not most, people (possibly including Wilson himself) are willing to say that the ascriptional practice is mistaken and hold on to the general belief.

9. The possibility of many equilibrium points: indeterminacy

One hopes, of course, that a systematic investigation of our ascriptional practices and the beliefs we associate with our semantic concepts will converge towards a stable interpretation of them. However, there is no guarantee that it will be possible to find any such convergence, or, even if it is, that there will be a *single* definitive way for the various elements to converge. Just as there may not be a single accessible equilibrium between the putative extensions and general characterizations of non-semantic terms such as, say, our old use of "fish" or "mass" (the former seeming to have at least two accessible equilibria, the latter quite possibly having no accessible equilibria),[76] there may be no single accessible equilibria between the putative extension and general characterizations for any of our semantic terms either. In particular, the answer to the question of whether, when certain bits of our practice conflict with our general beliefs, we should revise our beliefs or the practice itself depends upon our attitude towards the centrality of the attributions or beliefs in question. For instance, one's attitude as to whether or not whales are fish depends upon how central one takes the assumption that fish are a natural kind to be. If one takes it to be fairly central (as scientists are inclined to do), then it will seem 'obvious' that the best equilibrium will involve giving up the practice of calling whales fish. On the other hand, if one takes the assumption about natural kinds to be comparatively unimportant (as fishermen are inclined to do), then it may seem equally clear that one should reach an equilibrium in which whales *are* fish. In much the same way, our attitudes towards the comparative centrality of various semantic intuitions will be influenced by whether we are

interested in, say, cognitive science (naturalistic intuitions about meaning presupposed by the program will, of course, be fairly central) or epistemology (reductive/internalist intuitions may seem to be of little concern if they commit one to skepticism etc.). When such disputes occur, there may be no more reason to think that one of the competing semantic theorists is 'deeply confused about the nature of meaning' than there is to think that one of either the scientist or the fisherman must be 'deeply confused about the nature of fish.' Still, people who accept one equilibrium will have a tendency to view those who adopt another equilibrium as confused.[77] When such accusations are not made and each party accepts the usage of the other as a genuine alternative, there is still a tendency to treat one's own use as the norm. Scientists will contrast what fish are, 'strictly speaking,' with "fish" in its 'vulgar' sense, while laymen will contrast the 'normal' use of "fish" with its 'technical' use.

9.1 'Quinean' and 'Kuhnian' indeterminacy

Theorists who assign different weights to various general presuppositions about meaning and content may simply have 'incommensurable' views about how a coherent position should be reached, and this may produce genuine indeterminacy about how the contents of our thoughts and utterances should be properly individuated. That is to say, different interpreters, both fully aware of all the physical facts, might still legitimately disagree about how to interpret a given speaker. This possibility brings to mind, of course, Quine's thesis of the indeterminacy of translation, which he characterized as follows:

> Manuals for translating one language into another can be set up in divergent ways, all compatible with the totality of speech dispositions, yet incompatible with each other. In countless places they will diverge in giving, as their respective translations of a sentence of the one language, sentences of the other language which stand to each other in no plausible sense of equivalence however loose.[78]

Quine found translation to be indeterminate because all the relevant evidence underdetermined what the correct interpretations of a speaker's words were. The position

---

[77] As Ishmael's former shipmates put it, the scientist's reasons were "altogether insufficient," and, indeed, "humbug."
[78] Quine 1960, 27.

outlined above suggests that translation may (also) be indeterminate because what counts as the relevant evidence is *itself* indeterminate. Different interpreters, each fully informed about the facts about word usage in past, present, future and even counterfactual situations, could still disagree as to how these facts should be interpreted because they had different preconceptions about meaning and thus about what sort of evidence was relevant to an assignment of meaning, what the criteria for change of meaning were, etc.[79]

For instance, if two interpreters had decided to come up with manuals for my early use of "vegetable," one could quite easily imagine the following conflict emerging: the first interprets me as having the false belief that mushrooms are vegetables, while the second interprets me as meaning something different by "vegetable" than most of my peers. The first will interpret me as having meant the same thing by "vegetable" during my entire life, while the second will suppose that the meaning of this term changed for me when I stopped applying it to mushrooms. The first may give primary importance to being faithful to my own conception of what was happening (I took myself to have had a false belief about

---

[79] Some (see, Hylton 1982, 180) have argued that something of this sort was what Quine had in mind with his indeterminacy thesis, but Robert Kirk has argued that treating indeterminacy this way trivializes it:

> Exactness of literal equivalence by no means entails exactness of cultural equivalence. But .… Quinean indeterminacy would be a trivial thing if it depended on the fact that the constructors of different translation manuals may have divergent purposes and priorities. Not only would it be trivial, it would be quite useless for Quine because it would be consistent with the most extreme Platonism about meaning. For Platonists can easily agree that when translators have different purposes they will want to associate sentences with different Platonic meanings. If we are not to misrepresent the indeterminacy thesis, therefore, we must assume our translators share the same purposes and priorities, whatever these may be.… it seems best to assume that these purposes and priorities are scientific. The translators are to be supposed to be aiming at exact literal synonymy -- whatever that means exactly. (Kirk 1986, 73.)

> Kirk's criticism seems, however, to miss the mark, and his claim that indeterminacy stemming from interpreters having different purposes and priorities "is entirely consistent with extreme Platonist theories of meaning, and has no philosophical interest" seems to be based upon trivializing what these different purposes and priorities could be. Admittedly, if the differences all resulted because one interpreter was interested in preserving the "cultural" meaning of what is said while the other was interested in the "literal" meaning, the thesis would be of comparatively little interest. However, in the cases described above, all parties concerned are looking for the 'literal meaning' -- it's just that their "different criteria for the adequacy of a translation manual" cause them to disagree about what is relevant for determining the *literal* meaning. Kirk claims that "the translators are supposed to be aiming at exact literal synonymy -- whatever that means exactly," but whether there is anything which we exactly mean by "literal synonymy" is precisely what is at issue, and the advocate of Kuhnian indeterminacy may argue that there is nothing which we mean exactly by such a term.

mushrooms), while the second puts greater weight on, say, closely tying what I mean by my terms to those things I causally interact with. The first favors interpretations that attribute shared meanings to members of the same linguistic community, the second favors interpretations that maximize the amount of true beliefs on the interpretee's part. The two theorists may each see the value of the other's desiderata, but disagree over the comparative weight that should be placed upon them.[80]

That there could be such conflicts between interpreters should be clear to anyone with even a passing acquaintance with much recent work in the philosophy of language. Not only is the example structurally similar to that of Bert's use of "arthritis" (with Burge playing the role of the first interpreter, and, say, Davidson playing the role of the second),[81] but other examples of how philosophers differ over what criteria an interpretation should satisfy are also not hard to find. To mention just one such case, some philosophers would more or less discount any interpretation that did not seem as if it could be given some sort of naturalistic grounding,[82] while others would find this proposed criterion not only dispensable, but also in conflict with fairly central beliefs about the normative nature of semantic concepts. That this sort of conflict could produce 'incommensurable' commitments to divergent manuals has considerable plausibility.

There is little reason not to understand such disputes as instances of indeterminacy.[83] What Quine wanted, after all, was a case where "two translators might develop independent manuals of translation, both of them compatible with all speech behavior, and yet one manual would offer translations that the other translator would reject."[84] Davidson's and

---

[80] It should also be noted that this type of indeterminacy does not seem to conflict with the apparent determinacy we have from the 'first person point of view.' While I may be certain that I *never* meant "undetached rabbit part" by "rabbit," I have no such firmness of faith about what I meant by "vegetable" (indeed, I've changed my mind about it on more than one occasion).

[81] See, for instance, Burge 1979, Davidson 1987, 449, and Davidson 1994a, 6.

[82] For a recent expression of this view, see Devitt 1994.

[83] Though, of course, there is no guarantee that they will be. It is certainly *possible* that, say, Davidson and Burge have both worked out stable equilibrium points for our semantic terms, but whether this is, *in fact*, the case is something that will have to be investigated in what follows.

[84] Quine 1979, 167.

Burge's manuals clearly seem to fit the bill,[85] and there is no reason to think that any further utterances on Bert's part would settle the issue between them.[86] When the layman's use disagrees with that of the expert, Burge treats it as a genuine disagreement, while Davidson treats it as a verbal one.[87]

However, while the considerations raised above may suggest that examples of (and a motivation for) indeterminacy may not be difficult to find, it should be clear that the examples and motivation involved will be significantly different from Quine's. Indeed, one might draw a distinction between 'Quinean' and 'Kuhnian' indeterminacy. The former assumes agreement upon what counts as the relevant evidence and how it should be weighted and argues that such agreed upon evidence supports a number of conflicting theories equally well.[88] The latter is structurally similar to how our choice of scientific theories can be underdetermined due to 'incommensurable' differences between theorists about what counts as relevant evidence etc., and about how such evidence should be weighted.

The two types of indeterminacy need not be mutually exclusive. However, while there is little in Quine that positively rules out the possibility of Khunian indeterminacy, his positive views on the constraints on translation and his apparent belief that we couldn't find any actual instances of indeterminacy suggest that he does not have anything like Kuhnian indeterminacy in mind.

[85] Furthermore, one might want to claim that in the case of disputes about meaning of the sort Burge and Davidson would find themselves in, "either manual could be useful, but as to which was right and which was wrong there was no fact of the matter." (Quine 1979, 167.)

[86] This might strike some as a controversial claim, but unless Bert is unusually reflective about semantic issues pertaining to his own language, this will probably be true. Of course Bert could always *come* to be reflective in this way, but it would be open to either disputant to claim that such reflection had caused Bert to change what he meant and thus had no bearing on the interpretation of the original utterance.

[87] This sort of disagreement fits well, of course, with the point in "Two Dogmas" about there not being a firm distinction between differences in belief and differences in meaning once we give up the analytic/synthetic distinction. (See Quine 1953.) It is also in keeping with Davidson's claim that "when others think differently from us, no general principle … can force us to decide that our difference lies in our beliefs rather than our concepts." (Davidson 1974b, 197.)

[88] See Kuhn 1973, 326. [89] Indeed, Quine seems to have resigned himself to our being unable to discover any examples of the indeterminacy of translation proper, claiming instead that "full holophrastic indeterminacy of translation draws too broadly on language to admit of factual illustration." (Quine 1990, 50-51.)

The nature of Khunian indeterminacy also makes it clear why Quine himself had so much trouble coming up with an instance of the indeterminacy of translation proper.[89] Conflicting translation manuals, like conflicting accounts of the world's physical structure, are likely to be created by people working with conflicting theoretical or methodological commitments.[90] These commitments cause them to view the phenomena in different ways, as more or less important, etc. Quine, however, for all his talk about the rival manuals, never seemed to doubt that there was a single true 'theory of meaning' (namely, some version of the type of theory one found in *Word and Object*) which settled all the questions about what sort of evidence should be taken into account and what significance it should be given.[91] As a result, Quine's own methodology implicitly denies that the Kuhnian type of indeterminacy holds, and thus makes it extremely difficult for him to find an example for his own thesis.

9.2 Conventionalism

As discussed previously, when scientists discovered that they could not count whales as fish fishermen were not impressed. Nevertheless, while the two parties may have had some inclination to treat the others as 'confused about fish,' it seems quite clear that both ways of categorizing fish were recognizable extensions of the original "fish" practice, and the term could quite easily come to have two meanings (an 'everyday' and a 'technical' sense). If this option is available, then any dispute between the scientists and the fishermen would be a purely verbal one. Each pick out an equilibrium accessible from the previous

---

[90] This might also help explain why, after two decades of appearing to swallow the, on the face of it, incredibly implausible claim that there was no fact of the matter of whether we meant *rabbit* or *rabbit-part* by "rabbit," philosophers of language such as Davidson have so easily found themselves treating the question of whether content was individuated individualistically or not as a factual one. Gavagai-driven disagreements can arise between people who share the same general methodology, and are thus unable to discount their rival's manual. When the two interpreters do not agree on methodology, it is much easier for each interpreter to see the other's manual as being the product of 'conceptual confusion,' 'insensitivity to the real nature of psychological explanation,' 'lack of theoretical sophistication,' etc.

[91] "In psychology one may or may not be a behaviorist, but in linguistics one has no choice." (Quine 1987, 5.)

practice. It is simply a matter of convention which of the two we choose. Indeed, we might be able to freely switch from one to the other as we choose (as we seem to do with "weight," sometimes using it in its technical sense, and other times using it as more or less equivalent to "mass").

In much the same way, the viability of treating disputes between, say, Davidson and Burge as instances of indeterminacy may lead to a position where it is just a matter of convention how we should individuate content, with the consequence that different groups can do so in whichever way best suits their purposes. Just as the fisherman can work with a 'functional' conception of fish, while the scientist works with a 'biological' one, the computational linguist could work with an internalist conception of content, while the epistemologist works with an externalist one. Or, once again, the same person might use one while we are doing epistemology, another while doing AI, etc.[92]

However, the fact that our pre-theoretical intuitions about meaning pull us in many different directions by no means guarantees that there will be equilibria corresponding to each of these directions, and the fact that it may initially seem that there will be a stable equilibrium attached to some understanding of meaning by no means guarantees that there will be such an equilibrium.[93] Just as there is no guarantee that *only* one of them will be accessible, it is an open question whether, say, there really are *either* individualistic or non-individualistic equilibria.

---

[92] Perhaps forcing us to accept Lycan's "double indexical theory of meaning" ("MEANING =$_{def}$ Whatever aspect of linguistic activity happens to interest *me now*.") In any case, his claim that most accounts of meaning involve "disguised value judgments" (Lycan 1984, 272) would seem, on the model suggested above, to be correct.

[93] As is the case with the 'biological' conception of fish (assuming that biological kinds are picked out in evolutionary terms).

10. The possibility of no equilibrium points: bifurcationism

This brings us to the unpleasant possibility that there may not only be not many, but, in fact, no equilibria at all. There may be no way of making our general characterizations and putative extensions cohere that does not involve giving up more of either set than we are willing to. We usually suppose that there will be at least *some* coherent way to systematize our linguistic practices (i.e., we not only assume that there will be *at most* one equilibrium point, we assume that there will be *at least* one as well), but this presupposition can turn out to be mistaken.[94]

In such cases there will be a temptation to split the term (as was done with "weight" into "mass" and "weight," and then "mass" into "rest mass" and "relativistic mass") into two terms for which equilibria can be reached which together allow us to keep most of our original general characterizations and putative extensions. Those who make a division between 'narrow' and 'wide' content are effectively accepting such a view.[95] They conclude that we must split terms like "meaning" and "content" the way we had to split "mass" because they believe that there is no coherent way of systematizing the intuitions and ascriptional practices associated with our semantic terminology. A fairly clear expression of the motivation for such a view can be found in McGinn:

> Our intuitive conception of belief content combines two separable components answering to two distinct interests we have in ascriptions of belief. One component consists in mode of representation of things in the world; the other concerns itself with properly semantic relations between such representations and the things represented. I want to suggest that the former component is constitutive of the causal-explanatory role of belief, while the latter is bound up in our taking beliefs as bearers of truth .... We get different and potentially conflicting standards of individuation -- and hence different conceptions of what a belief essentially is -- according as we concentrate on one or the other component of content. The tendency of discussions of belief is, I think, to allow one component to eclipse the other, thus producing needless conundrums and a distorted picture of the nature of belief.[96]

---

[94] For doubts about the ability of linguistic phenomena to accommodate such a presupposition, see, for instance, Austin 1956, 151, Wilson 1982, 552.[95] See, for instance, Putnam 1981a and McGinn 1982, Loar 1985.

[96] McGinn 1982, 210-11. While McGinn expresses the motivation for the view very well, his view may not be a case of the view itself, since McGinn's official line seems to be that while "These components and the concerns they reflect are distinct and independent -- total content supervenes on both taken together." (210.)   Still, it is remains the case that, with McGinn, the explanatory work is being done by two independent components, so while there is, technically, just one notion of content for him, there are still two different things explaining behavior and being evaluated as true or false.

McGinn may be a little complacent about how easy it is to give up what he asks us to give up here, but a position such as his might ultimately turn out to be unavoidable. Accepting such "bifurcationism" requires, of course, denying that what it is that is externally determined and evaluated as true or false also plays a part in "commonsense psychological explanation."[97] However, while this 'unity' thesis has a strong *prima facie* plausibility, it need not be completely nonnegotiable.[98] It might be something that must be given up since, as long as it remains in our general characterizations, it prevents us from reaching any equilibrium. An equilibrium with it would be preferable to one without, but one without might be preferable to none at all.

It should be noted, however, that not all resolutions involving postulating two notions of content need be 'bifurcationist' in a sense that would correspond to a denial of the unity thesis. (Some of these alternatives will be discussed in chapter 2.) The unity constraint is only violated when the two 'types of content' split up the explanatory work in the way in which McGinn suggests that they do.

11. Self-reflexive character of dispute when applied to meaning

The dialectic is in many ways more interesting with semantic terminology because the outcome of the dialectic affects how we understand the dialectic itself. After all, the process of bringing the putative extensions and general characterizations for a term like "vegetable" into line does not, in itself, provide a characterization of what is going on in that process. On the other hand, when we become clear about our semantic terms we come to understand the dialectical process as one of either changing what we meant, coming to understand what we always meant, or coming to mean anything at all. The process of bringing the putative extensions and general characterizations of our semantic terms into line will itself provide us with an understanding of what is involved in that very process.

---

[97] Bilgrami 1992, 3, 15.

[98] Though it has occasionally been treated as such. For instance, Bilgrami finds the unity thesis central enough that he is willing to preserve it by "denying not only that meaning is social or normative in any interesting sense, but also by severing all of the standard associations of meaning and content from the official notion of truth-conditions." (Bilgrami 1992, vii.)

This self-referential quality can be seen by considering two alternative possible equilibria. The first, individualistic, alternative treats what a subject means by a term as constituted by his dispositions to apply the term and the beliefs he associates with it. Someone who thinks whales are fish is understood as simply meaning something different by "fish" than someone who does not (even if this belief does not mesh well with other beliefs of his that also help constitute what he means, such as that "fish" picks out a natural kind and that he means the same thing by "fish" as his compatriots). The second, non-individualistic, alternative treats what the subject means as constituted by whatever the (projected) equilibrium (if any) is for the practices that the society as a whole associates with the term.[99] Someone who thinks whales are fish simply has a false belief about fish (since that belief is one that would have to be dropped in order to reach an equilibrium).

If the equilibrium for our semantic terms were reached at a comparatively individualistic stage, then the dialectic itself would have to be viewed as a history of the gradual changes in what we meant by our semantic terms (or, at best, the history of how what we meant by our semantic terms became more definite over time). If the equilibrium state were individualistic, then someone who took a resolutely non-individualistic line on semantic terms would have to be understood as meaning something different by their semantic terms than the individualist, and thus could not, by the individualist's lights, be understood as having false beliefs about the very thing (namely meanings) that the individualist has true beliefs about (though he could, like the person who believed both that whales were fish and that "fish" picked out a natural kind, be criticized for associating an internally inconsistent set of beliefs and practices with his term).[100] On the other hand, if the dialectic ended up at

---

[99] These two options are not, of course, meant to be exhaustive. In particular, there could be an individualistic alternative that identified meaning with the equilibrium of the individual's beliefs and practices, not the beliefs and practices themselves, and there could be a non-individualistic alternative that identified meaning with the socially accepted beliefs and practices, not the equilibrium these beliefs and practices suggested.

[100] So the individualist could not even say the non-individualist claim that Bert meant *arthritis* by "arthritis" was false, since he must understand the non-individualist as meaning something different by "meaning," "reference" etc.

a very non-individualistic stage, then it might understand itself as the history of how we came to become clear about what we had meant by our semantic terms all along -- the individualist would be understood as simply mistaken about the very thing the non-individualist has come to have a correct conception of.[101]

Things are more complicated if we come to recognize that there can be more than one equilibrium (say, one for both of the individualist and non-individualist described above). In such a case, the individualist can understand himself as having correct beliefs associated with his semantic terms, even if it is the case the non-individualists beliefs about his semantic terms may be correct as well. Even if everyone else reaches a non-individualistic equilibrium, if the individualist can find an equilibrium for his semantic terms, then he can understand himself as correct. The non-individualist, however, will not only be unable to characterize the individualist as incorrect, but he will also be uncertain whether his own semantic beliefs are correct. Even if there is a non-individualistic equilibrium available for our semantic terms, the non-individualist can only have correct semantic beliefs associated with that equilibrium if that equilibrium is reached by the other members of his community. If everyone else reaches an individualistic equilibrium, then the non-individualist's non-individualism commits him to interpreting himself individualistically.

Finally, there is the case of there being no equilibria at all for our semantic terms, and so, no equilibrium corresponding to either the individualist's or the non-individualist's positions. In such a case, both will have to understand themselves as being, at least to a certain extent, mistaken about their semantic terms. If the lack of equilibria leads to a type of eliminativism, they will both have to understand themselves as very much mistaken, with all of their semantic beliefs being false. If the lack of equilibria leads to a type of bifurcationism, then each may be able to salvage some of their beliefs as true. For instance the bifurcationism might result in a non-individualistic 'wide content' and an individualistic

---

[101] As a result, while advocates of the second way of describing the dialectic can hope to show that their opponents are wrong, advocates of the former can only hope to show that, while their opponents were not wrong about 'the meaning of "meaning,"' they should (because of the place of equilibrium as a regulative ideal) move to a more individualistic conception.

'narrow content' with each theorist being able to understand their semantic beliefs as true about one of the two types of content and false about the other.

Discovering, then, how many (if any) equilibria there are for our semantic terms is of considerable importance. We can hope, if not expect, that our intuitions and practices will converge towards a single theory, and finding such a single equilibrium would obviously be of considerable theoretical interest. Finding more, or fewer, than one equilibrium, while in some sense less satisfactory, would be, of course, no less significant. If there are equilibria corresponding to, say, both Davidson's and Burge's positions, then debates between the defenders of these two positions remain, in some sense, futile, and what we are faced with are simply instances of indeterminacy. The possibility of there being no or multiple equilibria thus has significant methodological significance for debates about semantic issues, since the fact that there may not be a single equilibrium entails that showing that there is something wrong or 'incoherent' about alternative positions is neither a necessary nor a sufficient condition for establishing the adequacy of one's own position. It is not necessary, since the possibility of there being multiple equilibria allows that one's position can be adequate in spite of the adequacy of some alternative; it is not sufficient since the possibility of there being no equilibria makes it the case that the ultimate unworkability of all other positions does not entail the acceptability of the final alternative.

12. Conclusion

We have argued, then, that much of the normativity associated with language use arises from the tensions between the putative extensions and general characterizations associated with our terms. Both information-based and inferential role semantics were presented as unsuccessful attempts to explain semantic norms in terms of just one of these two elements that illustrate how the more dynamic relation between the two factors was necessary. It was further argued that a term's actual extension could be understood as what one gets if one can find a "equilibrium" between the general characterizations associated with a term

and its putative extension. However, when semantic norms are understood this way, a number of problems come to the fore. There is no guarantee that there will an equilibrium to be found for the practices associated with a term, and if one can be found, there is no guarantee that there will be *just* one. Settling, in such cases, whether what a particular term means remains indeterminate until an equilibrium is actually reached for it, or whether reaching an equilibrium makes manifest what the term meant all along turns out to involve getting a proper characterization of our semantic terms. This in turn involves finding an equilibrium for the interplay between their various putative extensions and general characterizations. A number of general beliefs about meaning were discussed, and a number of ascription types brought to our attention by Kripke, Putnam, Burge and Wilson were presented as providing evidence for the putative extensions of these semantic terms. Within this framework, indeterminacy was understood as the product of our being able to find more than one equilibrium for our semantic terms while bifurcationism was understood as the product of our being unable to find any such equilibria. Finally, the way finding an equilibrium for our semantic terms determines how we understand that very process was discussed. The next three chapters will involve discovering just how much of the putative extension made manifest in these ascriptions can be incorporateable into an equilibrium for our semantic terms. It will be suggested that there is an equilibrium that incorporates not only the ascriptional practices brought to our attention by Kripke, Putnam and Burge, but also those discussed by Wilson.

Chapter Two: The Expressive Constraint

1. Introduction

The previous chapter presented us with the task of finding a consistent equilibrium for our intuitions about meaning and content and left us with the disturbing prospect that there might be no such equilibrium for these terms. Those who think that no acceptable equilibrium can be found for all our intuitions in these areas often advocate a type of bifurcationism ('wide' vs. 'narrow' content etc.).[102] There is, however, another way of splitting the subject matter of one's semantic theory in order to make finding an equilibrium easier -- a way that might not strictly be considered bifurcationist. In particular, finding an equilibrium for our semantic terms might be easier if we make a clear distinction between linguistic and belief content.

There are a number of strong intuitions tied to the content of what we *believe* (relating to the explanation of behavior, first person authority, supervenience, etc.) that we do not feel (or at least do not feel as strongly) with respect to the content of what we *say*. In much the same way, there are a number of strong intuitions about what we mean by our words (meaning is conventional, public, shared, etc.) that are not felt as strongly about the contents of our beliefs. Because of this, if we were to look for separate equilibria for belief and linguistic content, the resulting two tasks would seem to be considerably easier than the task of finding a single equilibrium that worked for both. Indeed, much of the conflict that exists at the level of our general beliefs about meaning and content exists between the beliefs corresponding to the two proposed types of content.

However, there are general beliefs and ascriptional practices pertaining to the relation between belief and linguistic content as well. As a result, trying to find separate equilibria for belief and linguistic content may be more difficult than hoped. In

---

[102] For a more extended discussion of this, see section ten of chapter one.

particular, separating the two types of content would involve giving up what will be here referred to as the "expressive constraint," namely:

> When a speaker expresses a belief by sincerely uttering a sentence, the utterance and the belief have the same content.[103]

That is to say, when we make a sincere assertion, the content of that assertion is the same as the content of the belief that it is an expression of. The expressive constraint is not supposed to be true because it just so happens that the independently determined contents of our thoughts and utterances correspond, but rather because whatever determines the content of one thereby determines the content of the other. That is to say, the expressive constraint is not meant to be a mere empirical generalization, it is supposed to capture an 'internal' connection between the contents of our thoughts and utterances. The constraint itself involves no claim to priority on the part of either thought or language. As will be discussed in section four, both accounts that explain thought in terms of language and accounts that explain language in terms of thought can satisfy the constraint. The term "expression" and its cognates will henceforth used to correspond to this more 'internal' relation between thought and language.[104]

Splitting thought and linguistic content this way allows one to divide the subject matter of one's theory without some of the conceptual costs associated with many recent bifurcationist approaches. Such a split would not, for instance, violate Bilgrami's "unity constraint" (the requirement that the same contents both be evaluated as true or false and explain behavior),[105] since the belief contents will play this unified role. Linguistic contents, on the other hand, while they would still be evaluable as true or false, would play no role in the explanation of behavior. Linguistic contents would have no role in explaining our behavior other than providing a comparatively reliable source of evidence

---

[103] Forms of the expressive constraint can be found in Searle 1983, 164, McGinn 1982, 217, Bilgrami 1992, 1. Though, as will be discussed in 3.1, an implicit commitment to it is extremely widespread.

[104] There is a less loaded sense of "expression" (in which A expresses B if A is caused by B or provides some evidence for B) in which an utterance could be said to be an expression of a belief even if the two did not have the same content, but this less loaded sense will not be used here.

[105] See Bilgrami 1992.

for the speaker's beliefs. Furthermore, the thought/linguistic content split corresponds to an intuitive distinction between thought and language that has no analog in the narrow/wide content split. We occasionally do feel as if the content of what we said failed to capture what we were thinking, while we have no such experience of any split between the 'narrow' and 'wide' contents of our thoughts and utterances.

Just as someone can give up the expressive constraint without having to distinguish narrow and wide content, a theorist who posits a narrow/wide content split need not violate the expressive constraint. Both the narrow and wide contents could be found, and match expressively, in both the domains of thought and language.[106] Satisfying either the unity or the expressive constraint is thus neither a necessary nor a sufficient condition for satisfying the other.

2 Reasons for rejecting the expressive constraint

If the expressive constraint prevents one from using the 'divide and conquer' strategy mentioned above, what reasons could someone who wanted to use such an approach (hereafter "the anti-expressivist") provide for giving up the expressive constraint? Three possible reasons for giving up the constraint will be discussed in this section: (a) that it stands in the way of our finding equilibria for our semantic terms, (b) that it is really best understood as an empirical generalization, and (c) that its rejection is presupposed by the familiar speaker's reference/semantic reference distinction. None, I will argue, is ultimately convincing.

2.1 Easier to find equilibria without it

It seems undeniably true that, if we were in no way committed to the expressive constraint, it would be easier to find separate equilibria for belief and utterance content. Indeed, the anti-expressivist might argue that no equilibrium can be found that includes

---

[106] See, for instance, McGinn 1982 for an account that relies heavily on the narrow/wide split without violating (indeed, explicitly endorsing) the expressive constraint.

the constraint, but that separate equilibria can be found for both thought and linguistic content without it. If this were true, then it would seem to follow that, if we *could* give up the expressive constraint, we *should* give it up. Previous attempts to find an equilibrium might suggest that we can't find one that incorporates the constraint. As a result, it seems incumbent upon the defender of the expressive constraint either to show that such an equilibrium can be found with it, or to show that there could be no acceptable equilibrium for either belief or utterance content that did not incorporate some form of the expressive constraint.

However, given its *prima facie* plausibility, one only has reason to give it up if one can't find an equilibrium that includes it, and can find equilibria without it. This chapter argues that neither of these requirements are ultimately satisfied. The claim that one can't find an equilibrium without it (that is, that the expressive constraint captures non-negotiable aspects of our conception of thought and language) is discussed in section 3 of this chapter. That the constraint does not prevent one from finding an equilibrium is argued in section 4 of this chapter and in succeeding chapters. The first line of attack is thus inconclusive, it just sets the stage for the rest.

2.2 The constraint is best understood as an empirical generalization

A second line of attack on the expressive constraint argues that it involves treating what is merely an empirical generalization as a conceptual truth. The anti-expressivist can argue that, while the contents of our words *generally* reflect the contents of our thoughts, this is not because of the sort of 'internal connection' between the two that the expressive constraint requires. The anti-expressivist will instead suggest that the two contents generally correspond because, by and large, we know what we are thinking, we know what our words mean, and we make an effort to see that the sentences we utter correspond to the content of our thoughts. According to this line of thought, there is no 'internal' connection between the contents of our thoughts and utterances, but rather a

purely contingent one that can break down (for instance, when we make slips of the tongue or have an imperfect mastery of our language). As Evans puts it "We are all familiar with cases in which, though carelessness or ignorance of the language, the speaker selects words unsuitable to his thoughts."[107]   Dummett spells out this apparently familiar thought in greater detail as the following:

> When an utterance is made, what the speaker *says* depends upon the meanings of his words in the common language; but, if he thereby expresses a belief, the content of that belief depends on his personal understanding of those words, and thus on his idiolect … In unhappy cases, therefore, his words, understood according to their meanings in the common language, may not be the best expressions of his belief, or may even misrepresent it.[108]

If Dummett and Evans are right about this, the expressive constraint, in spite of its intuitive plausibility when formulated, is not actually reflected in our ascriptional practices. The content of what we say is determined by the conventional meanings of those words, while the content of the thought we are trying to express is determined by our understanding of those words. When Archie Bunker says "We need a few laughs to break up the monogamy," we attribute to him the belief that a few laughs are needed to break up the *monotony*. The conventional meaning of what he says does not determine the belief we attribute to him (indeed, the humor of Bunker's utterances would seem to be lost if we merely attributed to him the beliefs corresponding to the conventional meaning of his sentences). More (much more) tendentiously, one could argue that when someone like Bert, who has not fully mastered the term "arthritis," says "I've got arthritis in my thigh," we attribute to him a true belief about *tharthritis* that does not correspond to the conventional meaning of what he said (he said something false, but the belief he was trying to express was true). We may be inclined to assent to the expressive constraint, but we don't unexceptionally follow it when we actually go about our ascriptions.

---

[107]   Evans 1982, 67-8. Evans, of course, relies on a rejection of the constraint when he claims that Russell's principle applies for individuating the contents of what we think, but not for individuating the contents of what we say.

[108]   Dummett 1991, 88. It is not entirely clear how this is consistent with his earlier discussion of judgment as the interiorization of the external act of assertion (see Dummett 1973, ch. 10), though perhaps he is relying on some distinction between saying and asserting or believing and judging.

The expressive constraint is satisfied for most of our utterances, but it need not be satisfied for all of them. The anti-expressivist can thus suggest that what the expressivist perceives as an internal connection is really just an empirical generalization. The anti-expressivist will thus argue that thought and utterance content correspond only when certain empirical preconditions are met. Consider, for instance, Kripke's "disquotation principle," which might seem to be something very much like the constraint: "A normal English speaker who is not reticent will be disposed to sincere reflective assent to '*p*' if and only if he believes that *p*."[109] However, it is noteworthy that the principle is here taken to be true of "normal" and "reflective" speakers, and these qualifications are elaborated by Kripke below:

> When we suppose that we are dealing with a normal speaker of English, we mean that … he uses the sentence to mean what a normal speaker of English would mean by it.… The qualification "on reflection" guards against the possibility that a speaker may, through careless inattention to the meaning or other momentary conceptual or linguistic confusion, assert something he does not really mean, or assent to a sentence in linguistic error.[110]

Of course, qualified in this way, the disquotational principle holds good neither of anyone who makes a slip of the tongue or malapropism nor of anyone who, like Bert, has an imperfect mastery of the language. If Kripke's disquotation principle really captured all there was to the constraint, there would be no reason to see it as requiring any 'internal' connection between thought and language.

However, this line of attack is not as devastating as it might at first seem, and there is a way in which the defender of the expressive constraint can accommodate such apparent counterexamples, and thus reestablish the plausibility of the purported internal connection that the expressive constraint requires. In particular, he can reject Dummett's initial premise that "when an utterance is made, what the speaker *says* depends upon the meanings of his words in the common language."[111] The content of what we say may

---

frequently correspond to the conventional meaning of the uttered sentence, but it need not always do so. We can preserve the connection between thought and utterance content by loosening the connection between the contents of our utterances and the contents conventionally associated with such utterances.

This loosening should not be viewed merely as an *ad hoc* attempt to preserve the expressive constraint in the face of ascriptional practices that seem incompatible with it. Indeed, such a break between conventional meanings and the content of what was said would follow directly, for instance, from a Gricean analysis of language.[112] The meanings of our utterances are tied directly to our communicative intentions, and the contents of these intentions are tied to the contents of the corresponding thoughts. The 'conventional' meanings of the sentences in question are determined by how the sentences are *typically* used, and, as a result, there is no reason to think that any *particular* use of the sentence must have the conventional meaning.[113] The claim that the content of what we say is determined by the conventional meaning of our words is, at best, controversial.

The most clear and insistent case for a separation between the content of what we say and the meanings conventional conventionally associated with our words is found in the work of Donald Davidson.[114] Davidson makes a plausible (but by no means uncontroversial) case for the claim that, with malapropisms and other slips of the tongue, the literal meaning of what the speaker said differs from the conventional meaning associated with the words he uttered.[115] Of course we may not be *required* to take

---

[113] This giving primacy to communicative intentions is no mere artifact of the Gricean program. It has also been argued that 'conventional meaning' seriously underdetermines what we intuitively take most of our utterances to mean, so an appeal to communicative intentions, and thus thought contents, seems inevitable when determining the content of our utterances. (See Searle 1979.)

[114] See especially Davidson 1982, 1986a and 1994a.

[115] Davidson goes on to claim, of course, that this point extends to speakers such as Bert who 'misapply' the words in their language, and thus that conventional meaning has no relevance in determining the content of what a speaker says or thinks. I think he seriously overstates his case on this point, and this will be discussed further in section 4.

Davidson's line, but showing that this possibility is at least open makes it easier to hold on to the expressive constraint if serious problems are found with giving it up.

2.3 Speaker reference and semantic reference

One might also try to support a distinction between thought and utterance content by claiming that it corresponds to the frequently made distinction between speaker's reference and semantic reference.[116] On such an account, the speaker's reference of our terms corresponds to our understanding of them, and thus our thought content, while the semantic reference corresponds to the conventional meaning, and thus utterance content. The expressive constraint would thus hold, in some sense, 'internally' between what we are thinking of and our speaker's reference, but only have a contingent empirical connection to the semantic referents of our terms. One could claim that, say, someone who believed that all stools were chairs would have "chair" mean *chair or stool* when we were specifying the content of their beliefs, but just *chair* when we were specifying the content of their utterances. One could then try to support this claim by arguing that the *speaker's* reference for his use of "chair" would include stools, but the *semantic* reference for his use of "chair" is still identical to the societal norm.

However, such an assimilation of these divergences from the societal norm with the speaker's reference/semantic reference distinction would be a mistake. First of all, instances of the classic speaker's vs. semantic reference phenomena can be generated *within* what supporters of this proposal would have to treat as cases of speaker's reference. For example, if the speaker (still calling stools "chairs") saw a large waste-paper basket in the distance and, mistaking it for a stool, said "the chair in the distance is awfully small," it should be clear that his speaker's reference is the wastepaper basket, not anything within the class of stools and chairs. Furthermore, speaker's reference is a pragmatic notion, which is supposed to be essentially context dependent (witness all the examples of speaker's reference in the literature). That is to say, the speaker's reference/semantic reference distinction is standardly taken to arise in cases where the intended reference can be demonstratively identified (hence its relation to the

---

[116] For a discussion of the speaker's-reference/semantic-reference distinction, see Kripke 1977.

referential/attributive distinction), and the utterance content/belief content split, if found at all, should be found in a much wider range of cases than this.   As a result, the speaker's reference/semantic reference distinction does not seem well suited to map onto the comparatively context independent divide between belief and utterance content suggested here. Finally, the speaker's reference/semantic reference divergences found in utterances such as "Mary's husband is kind to her" (said of someone who is not Mary's husband) are cases where it seems that there are *two* beliefs related to the speaker's utterance (the belief that Mary's husband is kind to her and the belief that that man is kind to her). Since there appear to be beliefs corresponding to the utterance understood both in terms of the speaker's references and its semantic references it does not seem promising to try to understand belief content merely in terms of the speaker's reference. As a result, even if we accept the speaker's reference/semantic reference distinction, it gives us no reason to think that there will be a corresponding distinction between thought and utterance content.

3. Reasons for keeping the constraint

We have now seen some of the more popular reasons for giving up the expressive constraint. What reason can we find for thinking that we need to *accept* the expressive constraint?   There is the obvious fact that, when stated, it seems to be true (indeed, it is often presented as something so obviously true that it doesn't need to be argued for).[117] Still, given how much easier finding an equilibrium would be if we weren't committed to it, an argument for accepting the expressive constraint will have to rely heavily on the unacceptability of the consequences of replacing the 'internal' connection between thought and language with a more contingent one. Fortunately for the defender of the expressive constraint, there are a number of such consequences.

---

[117] For instance, neither Searle nor Bilgrami argue for the versions of the constraint quoted in note 1. Still, one might claim that while it seems 'obvious' that something like the expressive constraint holds, the claim that what is obvious is the purported conceptual truth rather than the empirical generalization is much more contentious.

Among these reasons are (a) the fact that most philosophical accounts of meaning, while they do not go out of their way to defend the expressive constraint, do end up being committed to it; (b) giving up on the constraint undermines our traditional belief that we can learn what another is thinking from his words; (c) while our self-ascriptions are frequently taken to be authoritative, they might often be false if the expressive constraint were not satisfied, and (d) the type of conceptual connection between belief and assertion that is manifested in the paradoxical nature of 'Moore-sentences' such as "*P* but I don't believe it" cannot be made sense of if the expressive constraint is given up. All of these reasons will be discussed below.

3.1 Philosophical accounts of meaning

While only a few examples were given of philosophers who explicitly held it, a commitment to the expressive constraint is not an isolated quirk of philosophers such as Bilgrami and Searle. Most philosophical accounts of meaning involve some commitment to the expressive constraint. This does, not, of course, entail that the constraint must be accepted, but it does suggest that one can be quite naturally led to a commitment to it when one thinks about language. The fact that the constraint is entailed by most of these attempts to give a systematic account of thought and language suggests that the expressive constraint lies close to the heart of our presystematic semantic intuitions. Among the philosophical theories that involve a commitment to the constraint are the following.

3.11 Classical accounts

'Classical' accounts of meaning in which the meaning of a term was determined by the idea it was associated with clearly involve a commitment to the expressive constraint. Consider, for instance, the following passages from Hobbes and Locke:

> The generall use of Speech, is to transferre our Mentall Discourse, into Verbal; or the Trayne of our Thoughts, into a Trayne of Words; and that for two commodities; whereof one is, the Registring of the

> Consequences of our Thoughts; which being apt to slip out of our memory, and put us to a new labour, may again be recalled, by such words as they were marked by.[118]

> Man, though he had a great variety of thoughts, and such from which others as well as himself might receive profit and delight; yet they were all within his own breast, invisible and hidden from others, nor can of themselves be made to appear. The comfort and advantage of society not being to be had without communication of thoughts, it was necessary that man should find out some external sensible signs, whereof those invisible ideas, which his thoughts are made up of, might be made known to others .... Thus we may conceive of how *words* ... came to be made use of by men as the signs of their ideas; not by any natural connection that there is between particular articulate sounds and certain ideas ... but by voluntary imposition, whereby such a word is made arbitrarily the mark of such an idea.[119]

If speech merely "transforms our mental discourse into verbal," or if words are "arbitrarily made the mark" of our ideas by such "voluntary imposition," it should be clear that the content of our thoughts and utterances (being combinations of our ideas and words) will be the same. Since what I mean by my words is determined by the content of my thoughts, it will be impossible for the two to differ, and the expressive constraint will be satisfied.

Few people accept the classical account of meaning, though one might view generative semantics and at least some versions of the 'language of thought' hypothesis as an attempt to resuscitate some form of picture. Nevertheless, the general strategy of explaining the content of our utterances in terms of the contents of our thoughts can be found in much contemporary work.

This shouldn't be surprising. The thought that 'meaning' belongs intrinsically to our thoughts and comes only derivatively to our words (which, on their own, are just lifeless sounds or marks) can seem almost inescapable at times and, once this thought is in place, some form of the order of explanation posited in the classical picture will seem inevitable.[120]

---

[118] Hobbes, 1651, 101.
[119] Locke 1975, Book III, Chapter II. [120] This type of divide is stressed heavily in Searle 1992 when he contrasts the 'intrinsic' intentionality of thought and the 'derived' intentionality of language.

3.12 Gricean accounts[121]

Gricean accounts of meaning (or 'intention based semantics'), in which the meanings of our utterances are determined by our communicative intentions, also involve a fairly clear commitment to the expressive constraint. On such accounts, a speaker's meaning P by an utterance of "X" is explained in terms of his intention to communicate that P by his utterance of "X." In particular, on such accounts, speaker S is taken to mean P by uttering "X" to A if S satisfies roughly the following requirement:

S uttered "X" intending that:

    a. A be informed that P.
    b. A recognize S's intention (a).
    c. Intention (a) is satisfied in part thought the satisfaction of intention (b).[122]

Once this account of speaker meaning is in place, word meaning is explained in terms of practices or conventions of using words with a particular speaker meaning. For instance, "X" means P timelessly if there is a convention or practice of using "X" with the speaker's meaning P.

The Griceian story outlined above would, if true, entail that some form of the expressive constraint was true as well. On Gricean accounts, what we mean by our words is determined by what we intend to communicate, and so the belief contents associated with our communicative intentions will be reflected in the contents of our utterances. Griceans can thus understand the expressive constraint as a product of this connection between communicative intentions and utterance contents. As a result, if we reject the expressive constraint, we must give up on all such Gricean accounts as well.

Furthermore, while one might have doubts about the plausibility of Gricean accounts when understood as providing a framework for *reducing* linguistic meaning to communicative intentions,[123] understood as a non-reductive analysis of the relation

---

[121] See Grice 1957, 1968, 1987, Schiffer 1972, Bennett 1976, Appiah 1985, Sperber and Wilson 1986.
[122] Such simple formulations of the account are subject to a number of counterexamples (see, for instance, Grice 1969, 1982, Schiffer 1972, Searle 1970, Strawson 1964). The altered versions are considerably more complicated, but for present purposes are not relevantly different.
[123] For criticisms of such reductive ambitions, see Rosenberg 1974.

between meaning and intention, Gricean accounts (which are not restricted to purely 'linguistic' meaning) remain extremely intuitive. As a result, giving Gricean accounts (rather than the reductive ambitions associated with them) up is something we should be hesitant to do.

3.13 Davidsonian accounts

If one were to give up the expressive constraint it would seem that the projects of determining what a speaker means by his words and determining what a speaker believes could, in principle, be carried out independently of one another. Indeed, if what a speaker meant by his words was determined by, say, the linguistic conventions of his community, then it would seem as if one could determine what a speaker meant by his words without observing the speaker at all. Such a possibility is, however, anathema to writers within the Quinean tradition such as Davidson. For Davidson, the projects of discovering what a speaker means by his words and discovering what he believes are just two sides of the same coin. As Davidson puts it "we should think of meanings and beliefs as interrelated constructs of a single theory."[124] He later argues that "if we know he holds the sentence true and we know how to interpret it, then we can make a correct attribution of belief. Symmetrically, if we know what belief a sentence held true expresses, we know how to interpret it."[125] The Davidsonian understanding of belief and assertion in terms of holding sentences true quickly leads to a version of the expressive constraint. If I am taken to hold true the sentences I assert, and thus to believe them, then from the fact that I assert that *P* it would seem to follow that I believe that *P*. It is not surprising, then, that writers who try to reject the constraint try to drive a wedge between holding "P" true and believing that *P*. Their idea is that the first leads to the second only if one has mastered all the concepts involved in understanding "P." Such a commitment to the expressive constraint can be found in any writer who endorses the 'radical

---

[124] Davidson 1974, 146.
[125] Davidson 1975, 162.

interpretation' methodology. Of course the Davidsonian program is not without its critics,[126] but even those who have problems with the radical interpretation methodology itself do not criticize it for its endorsement of the expressive constraint.

3.14 'Sellarsian' accounts[127]'Sellarsian' accounts of thought and language also involve a commitment to the expressive constraint. On such accounts, "X believes that *P*" is just supposed to commit its utterer to X's commitment to *P* (without endorsing that commitment), while "*P*" commits the utterer to the truth of *P*. On such accounts it thus follows that "I believe that *P*" commits one to one's own commitment to *P*. If one sincerely asserts that "*P*", one thus aquires the commitment that will make one's self-ascription of the belief that *P* true, and thus make it true that one believes that *P*. As a result, if one sincerely asserts that *P*, then one must count as believing *P* as well, and the expressive constraint will be satisfied.

3.15 The descriptive theory of reference

Descriptive theories of reference can be characterized as follows:[128]

(1) To every name or designating expression 'X,' there corresponds a cluster of properties, namely the family of those properties @ such that *A* believes '@X.'

(2) One of the properties, or some conjointly, are believed by *A* to pick out some individual uniquely.

(3) If most, or a weighted most, of the @'s are satisfied by one unique object *y*, then *y* is the referent of '*X*.'

(4) If the vote yields no unique object, '*X*' does not refer.

On such an account, what our words refer to is determined by our beliefs.[129] As a result, the content of what we say will be so determined as well. If what our words referred to

---

126 See, for instance, Searle 1987.
127 See Sellars 1963, Brandom 1994.
128 This characterization follows Kripke 1972, 71.
129 Or at least this will be the case for the fragment of our language for which the description theory is true. Russell defends a description theory but thinks it does not hold for "logically proper names,"

was determined by the beliefs we associated with those words, then we would expect belief and utterance content to match. Consequently, the expressive constraint would be satisfied on any such theory.

3.2 Communication and belief attribution

If the expressive constraint were given up, we would have to conclude that the language we speak can be, at best, an imperfect medium for communicating our thoughts. This is because we would have no guarantee that the content of what we say would correspond to the content of the thought we were trying to express. With this connection broken, we could not be sure that we would be able to let someone know what we were thinking through our utterances, nor could we be sure that we could correctly determine what someone is thinking by listening to what she says. We wouldn't know what another is thinking, not because he may mean something different by his words than we do, but rather because, while we both mean the same thing by our words, the contents of our sayings will not correspond to those of our thoughts.

As a result, without the expressive constraint, if X says "P" and we go on to make the ascription "X believes that P," there is a very good chance that this ascription will be false. However, we usually judge such ascriptions to be true. As a result, giving up the constraint would force us to give up fairly well entrenched beliefs about the accuracy of our ascriptional practices. Unless it can be shown that giving up these beliefs is ultimately unavoidable, the fact that giving up the expressive constraint would force us to give up these beliefs as well gives up an at least *prima facie* reason for not giving up the constraint.Finally, it should be noted here that while the thought that someone's sincere utterances allow us to tell what they are thinking is an extremely appealing one, it is not always an easy one to account for. It is clear that giving up the expressive constraint entails giving up on this appealing thought, but it should not be thought that one can

---

Devitt actually seems to endorse a description theory for all terms other than natural kind terms and proper names. (See Russell 1905, 1918, Devitt & Sterelny 1987.)

account for the thought just by accepting the constraint. Indeed many theories that entail the constraint still have problems in this area (in particular, how "belief-theoretic" accounts of the expressive constraint still have this problem will be discussed in section 4.21). Its incorporating some form of the expressive constraint may be a necessary condition for an account's capturing our intuitions about understanding others, but it is not a sufficient one. As a result, considerations relating to our understanding of others will not only give us reason to accept the expressive constraint, but also give us reason to accept a particular account of it. (Such an account will be provided in section 4.3 of this chapter.)

3.3 Self-ascriptions

A related, and perhaps more serious, consequence of giving up the expressive constraint would be that one's *self*-ascriptions of belief will often turn out to be false. If one has a less than perfect mastery of the terms in one's language (which, of course, we all do) then, even without malapropisms and slips of the tongue, the content of our *belief* that *P* (which is determined by our *understanding* of the conventional meanings (call it *P1*)) will not correspond to the content of our *utterance* "*P*" (which is determined by the conventional meanings themselves (call it *P2*)). As a result, if one makes a self-ascription of the form "I believe that *P*," the content of the '*P*' in one's self-ascription is *P2*, but the content of one's belief is *P1*. One may not *believe P2* at all, so the self-ascription may turn out to be false. Dummett alludes to those "unhappy cases" where a speaker's words "may not be the best expression of his belief, or may even misrepresent it,"[130] but if the expressive constraint really is given up, these 'unhappy cases' may occur all too frequently. Indeed, they may even be the norm. Our self-ascriptions would not strictly be true whenever our own mastery of the language is less than complete.

---

[130] Dummett 1991, 88.

However, our reliance upon the so called "division of linguistic labor"[131] suggests that no one could have complete mastery of their own language. As a result, such slippage is inevitable for any term for which we rely on the division of linguistic labor.

One possible way to avoid this problem would be to adopt the 'Fregean' strategy of claiming that words in belief contexts do not have their normal referents. In this case, a word's meaning is determined by the standard linguistic conventions in 'normal' contexts (such as "arthritis is painful"), but by the speaker's own understanding of those conventions in 'propositional-attitude' contexts (such as "John thinks that arthritis is painful").[132] However, if these thought contents *can* be expressed linguistically, it isn't clear why they shouldn't *always* be. Why couldn't we attach a *sotto voce* "I believe that" at the beginning of each utterance? Once linguistic meaning is allowed to be determined by thought content for some of our utterances, it isn't clear why it shouldn't always be so. Furthermore, even if one were able to accept such a loss of 'semantic innocence,'[133] one would be left with the previously mentioned problem that belief ascriptions made of other people ("John believes that *P*") are even more likely to be false on such an account, since the truth of such ascriptions would be dependent upon John's and the ascriber's idiosyncratic understandings of *P* happening to coincide.

We take our self-ascriptions to be in pretty good shape unless we have some particularly psychological reason for their not being so (repression, denial, etc.). It is a deeply entrenched aspect of our reactions to the belief reports of others that we take someone who sincerely claims to believe that *P* to actually believe that *P*. Because of this, the fact that giving up the expressive constraint would make these self-ascriptions

---

[131] A reliance that Putnam takes to be universal and Burge takes to be a psychological necessity for human beings. (See Putnam 1975, Burge 1989.)

[132] The proposal here is 'Fregean' only in the sense that it suggests that words have a different reference in attitude contexts. The proposal here differs from standard Fregean ones in that the belief context does not pick out the mode of presentation of what is picked out in the plain contexts, but rather picks out something completely different. "Monogamy" and "monotony" have completely different references, and the second could hardly be considered the mode of presentation of the first.

[133] Davidson, 1968.

unreliable enough to make this deeply entrenched inferential practice unjustified provides one a good reason for holding on to it.

3.4 Knowledge of meaning

In a similar fashion, giving up the expressive constraint undermines our purported knowledge of what we mean by our own words. One can still *say* things like "My word 'arthritis' refers to arthritis" and count on such utterances being true, but thoughts of the form "my word "arthritis" refers to arthritis" need not be true, since what determines the content of one's arthritis-thoughts is not the same as what determines what one's word "arthritis" refers to. Once again, that such thoughts about what our words refer to are true is a deeply entrenched aspect of our thought about language. I take it as obvious that, for instance, my thoughts of the form "my word 'mosquito' refers to mosquitoes" are true. If the expressive constraint is given up, however, such 'obvious truths' could turn out to be false, so we have yet another reason for thinking that giving up the expressive constraint would force us to give up deeply entrenched aspects of our semantic practice.

3.5 Moore's paradox

Perhaps the most serious problem with giving up the expressive constraint is that it is intimately connected to the intuitions behind Moore's paradox. If one gives up the expressive constraint, one can no longer make sense of what is supposed to be so 'paradoxical' about the 'Moore sentences' of the type "*P* but I don't believe it." If linguistic and thought contents differ then one might frequently be entitled to make assertions such as "Positrons have mass, but I don't believe it." This is because, while one might know that the public truth/assertion conditions (whatever they may be) of "Positrons have mass" are satisfied, one might also know that (because of one's half-baked knowledge of physics) the content of *one's own* belief "positrons have mass" will not correspond to the content of the assertion made in the public language. (In much the

same way, a third party could truly claim that one sincerely asserted that positrons had mass but didn't believe it.)

The anti-expressivist might argue that the Moore sentences could be understood as making manifest some sort of pragmatic problem. It might be viewed as part of a cooperative maxim that one does one's best to know what one's words mean. A sincere assertion of a Moore sentence, even if true, would make manifest one's having failed to satisfy the cooperative principle, hence, the anti-expressivist would argue, their anomalous quality. Such an account is obviously unsatisfying. Even those who give pragmatic accounts of the Moore sentences tend to do so in terms of some kind of pragmatic *contradiction* not in terms of a mere violation of the conversational maxims. A Moore sentence is much more anomalous than a long-winded, irrelevant, or any other standard violation of the cooperative maxims. As a result, accounts of their anomalousness in terms of their violating such maxims seem unpromising.

Of course, the anti-expressivist may take this 'dissolution' of Moore's paradox to be an argument in favor of his position, but generally these considerations should give us an at least *prima facie* reason to think that we should hold on to the expressive constraint. The paradoxical nature of the Moore sentences is an expression of our commitment to the expressive constraint.

The anti-expressivist might try to come up with a 'Fregean' solution[134] to this problem as well. The "P" in the Moore-sentence may not itself be in a belief clause, but one could stipulate that if the pro-sentence ("it") is in a belief clause, then the sentence it is anaphorically dependent upon should be interpreted as if it were as well. This move seems fairly *ad hoc,* and would not, unfortunately, give an account of the paradoxical nature of sentences such as "Snow is white but I don't believe that snow is white" where the two occurrences of "snow is white" would require different interpretations on the Fregean reading. Furthermore, the Fregean/pro-sentential position would here seem to

---

[134] For the use of this term, see the previous section.

entail that if we responded to another's assertion by saying "I don't believe it," then the other speaker's assertion should be interpreted in terms of our thought contents. One cannot claim that the pro-sentence just creates a thought content analog of the sentence it is dependent upon (that is, it functions as a type of 'pronoun of laziness' inserting an independent element into the belief context), or the Moore-sentences would remain unparadoxically assertable. This is an extremely unattractive conclusion, since it seems just crazy to claim that the correct interpretation of the words in one speaker's assertion is determined by how such words are understood by whichever speaker says "I don't believe it," when faced with the assertion. As a result, the Fregean/pro-sentential account of the Moore-sentences seems to be of no help to the anti-expressivist.

Giving up the expressive constraint thus only seems possible at considerable conceptual cost, and it is worth looking at how we could accept the constraint, and explain its truth once it is accepted. The following section will consist of an attempt to do just that.

4. Explaining the constraint

4.1 Belief- and meaning- theoretic accounts of the constraint

If one accepts the expressive constraint, then one should try to find an explanation for the purported 'internal' connection between thought and utterance content. Perhaps the most obvious ways to explain this connection involve understanding on of either belief or utterance content in terms of the other. If one of either belief or utterance content determines the other, then it should be no surprise that there is an 'internal' connection between the two. Most theories about the relation between thought and language can be understood as taking one of these two forms.

We can call accounts that take thought contents as basic (and try to understand linguistic meaning in terms of the thoughts our words express) "belief-theoretic accounts of meaning." Such accounts focus, predictably enough, on the role of language in

thought. They tend to be individualistic, and tied to fairly robust[135] conceptions of self-knowledge, language mastery and behavioral explanation (causal). On the other hand, we can call accounts that take linguistic meaning in a public language as basic (and try to understand thought contents in terms of the meaning of the sentences we use to express them) "meaning-theoretic accounts of belief." Such accounts focus primarily on the role of language in communication. They tend to be non-individualistic, and tied more closely to issues relating to truth, communication, and more normative explanations/evaluations of behavior. Belief-theoretic accounts have the most intuitive pull when we focus on the contents of our thoughts, our knowledge of what we mean, etc., while meaning-theoretic accounts have the most intuitive pull when we focus on our ability to communicate, the content of what we say, the truth or falsity of our beliefs, etc.[136]

With the much discussed 'linguistic turn' in philosophy, and the resulting shift of focus from the content of our ideas to the meanings of our words, one might think that the meaning-theoretic approach would be characteristic of most contemporary theories. It is, however, by no means obvious that this is the case. The Gricean program is not the only belief-theoretic one around today.[137] Philosophers such as, among others, Chisholm, Chomsky, Fodor, Searle and Stalnaker all fall pretty clearly into the belief-theoretic camp.

---

[135] That is to say, our knowledge of what we mean is not understood simply as the fact that sentences like "by 'pig' I mean *pig*" will always be true, or as constituted by the fact that an interpreter must treat the interpretee as knowing what he means, etc. (For examples of such non-robust accounts, see Davidson 1987, Burge 1988 (both of which are discussed in chapter three).)

[136] Versions of the belief and meaning-theoretic accounts were defended by Chisholm and Sellars respectively in their famous "Correspondence on Intentionality" (Sellars and Chisholm 1958). [137] While Grice's own account of meaning, and most of those inspired by him, are best understood as versions of the belief-theoretic account, it should be noted that the general Gricean formulation does not entail any such commitment. (For a discussion of this, see section 3.12.)

'Interpretational theories'[138] with their purported aim of arriving at a theory of translation or truth for a speaker's language might seem like a better candidate for the meaning-theoretic camp, but they can slide remarkably easily into a belief-theoretic form as well.[139] The theories of meaning and belief are supposed to be "interrelated constructs,"[140] but the major constraint on the joint theory, charity, has more to do with what the speaker can be taken to believe than with what he can be taken to say.[141] Quine refuses to treat what the "native's" compatriots have to say as being at all relevant to how his words should be translated (other than, of course, its having played a causal role in shaping our present use). In much the same way, Davidson is reluctant to see social use as having anything to do with what we mean. To the extent to which such commitments are defensible at all, they are defensible from within a roughly belief-theoretic framework. Within a meaning-theoretic framework, such restrictions upon the relevant evidence (which are introduced with little defense) seem completely unmotivated.[142] While interpretational accounts might be developed in ways that were not belief-theoretic,[143] most of the actual theorists working within this paradigm have tended to understand interpretation this way. Meaning-theoretic accounts are considerably less common. Sellars[144] and Selarsians such as Brandom[145] may be viewed as working with

---

[138] I.e.: theories that give pride of place to the process of "radical" translation or interpretation, such as those found in writers such as Quine, Davidson, and Bilgrami.

[139] This tendency is most explicit and unapologetic in Bilgrami 1992a.

[140] Davidson 1974, 146. [141] This shifting of the domain of charity from the domain of language to that of belief is implicit in Quine, and becomes increasingly explicit in writers such as Davidson, Dennett and Haugeland. Stich is self-consciously explicit about the shift when he claims that "the most illuminating reading of [Quine's discussion of charity] is to view it not only as a contribution to the theory of translation, but also and more fundamentally, as an attempt to set out some conditions constraining the intentional characterization of a speaker's beliefs." (Stich 1990, 31-4. See also Stich 1984 214-6.)

[142] Though Davidson has subsequently tried to defend them.

[143] And thus more in keeping with Davidson's claim that "Neither language nor thinking can be fully explained in terms of the other, and neither has conceptual priority." (Davidson, 1975 156.)

[144] See especially the letters in Sellars and Chisholm 1958.

[145] Brandom 1994.

such an account, and Tyler Burge[146] and the interpretation of Wittgenstein advocated by Kripke[147] are both frequently treated as being in the meaning-theoretic camp,[148] but the view has been surprisingly unpopular.

Indeed, meaning-theoretic accounts seem only to have come to the fore since the mid 70's when the work of Putnam and Kripke called into question whether belief-theoretic accounts really could give an adequate account of linguistic meaning (since belief-theoretic accounts were taken to be committed to the slogan "meaning is in the head"). Still, the meaning-theoretic account didn't follow immediately with Kripke and Putnam's work. Kripke said comparatively little about belief contents in *Naming and Necessity*, and Putnam's "The Meaning of 'Meaning'" (with its explicit divorce of what a speaker means by his words from his 'psychological state') could probably be better read as a rejection of the expressive constraint than an advocation of a meaning-theoretic view.[149] Still, once 'causal' and 'historical' accounts of reference were in play for our words, they quickly came to be viewed as adequate theories of thought content as well. This was especially clear with Burge, who explicitly directs his arguments towards the contents of our thoughts rather than (but not to the exclusion of) the meanings of our words.[150]

It seems then, that both belief- and meaning-theoretic accounts of the expressive constraint are possible, and, indeed, have been offered in the past. However, both such accounts face serious problems of their own, and these problems will be the topic of the next section.

4.2 Problems with the two accounts

---

[146] Especially Burge 1979.
[147] In Kripke 1982.
[148] Though I should note that this does not strike me as a charitable reading of either Kripke or Burge.
[149] The lingering commitment to psychological states still being 'in the head' is, of course, one of the aspects of that paper that Putnam quickly ceased to endorse.
[150] See especially Burge 1979, 1986a. Still, it is not clear that Burge's position can correctly be described as a meaning-theoretic one (see especially Burge 1986c).

Both belief- and meaning-theoretic accounts are reductive. Belief-theoretic accounts are committed to providing an account of what determines the content of the speaker's beliefs without making any reference to what the terms in his language mean. That is to say, they require that one be able to determine what any given person means by his words and believes independently of determining the *conventional* meanings of the words he uses. (This is obvious in Quine's methodology, where the speaker's interactions with other members of his community are explicitly ruled out as contamination of the translator's evidence.) In much the same way, meaning-theoretic accounts are committed to providing an account of what a speaker means that does not help itself to his beliefs. Each requires that one explain either the content of the speaker's thoughts or the content of his utterances without appealing to the other. The 'purity' that these theories strive for causes them to have a number of unintuitive consequences. In particular, it will be argued that belief-theoretic accounts drive too large a wedge between conventional meaning and belief content, while meaning-theoretic accounts fail to leave enough room for there to be such a wedge at all.

4.21 Problems with belief-theoretic accounts

Belief-theoretic accounts try to give an independent account of belief contents, and then assign these contents directly to the speaker's words. However, if belief contents are to be assigned independently of any reference to what a speaker's words mean in a public language, then one should expect the resulting belief and linguistic contents to be individualistic.[151] According to belief-theoretic accounts, speakers like Bert are best understood as meaning *tharthritis* rather than *arthritis* by "arthritis." The assignment of *tharthritis* seems to 'fit' his actual tendencies to apply the term and his 'arthritis'-beliefs. On the other hand, all of the reasons that might favor the assignment of *arthritis* involve

---

[151] Indeed, many theorists working within a belief-theoretic framework take on with it a commitment to individualism about belief content. This commitment is most explicit in Bilgrami's *Belief and Meaning*, and Davidson's occasional polemics against non-individualism (see for instance, Davidson 1986, 1994a) often stem from his taking a belief-theoretic conception of what his project involves.

reference to the meaning *conventionally* associated with "arthritis," and this is precisely the sort of information that the belief-theorist cannot allow to be relevant. The pros and cons of individualism will be discussed in more detail in the next chapter, but, suffice it to say, there are enough cons associated with it to make a commitment to it be undesirable.

It should, however, be noted here that, like theories that reject the expressive constraint, belief-theoretic accounts have the disadvantage of making our learning what another believes from his sincere utterances problematic. The content of a speaker's utterances may reflect the content of his beliefs, but there is no reason to think that the speaker will mean the same thing by his words as his interlocutor (though one should expect the two meanings to be similar). As a result, one's interlocutors will often not fully understand one's utterances, and thus they should not be expected to grasp the content of one's beliefs. A related problem will be, of course, that many of the belief-ascriptions we make about others will be false. When the doctor claims that Bert believes (falsely) that he has arthritis in his thigh, his ascription is false because Bert believes (correctly) that he has *tharthritis* in his thigh.[152]

4.22 Problems with meaning-theoretic accounts

Meaning-theoretic accounts try to give an independent explanation of linguistic meaning (that is, one that does not help itself to the contents of the speaker's beliefs) and then attribute to the speaker beliefs with the contents of the sentences that would be used to express them. Such accounts bring with them problems which, while very different from those that faced the belief-theoretic accounts, are no less serious.

While belief-theoretic accounts tend to be individualistic, meaning-theoretic accounts frequently go to the corresponding extreme of being *anti*-individualistic. That is to say, it's not just that the contents of the speaker's words and thoughts can be determined by

---

[152] As will be discussed in greater detail in the next chapter, this will also entail that Bert's ascriptions of belief to his *past* self will tend to be false.

factors that extend beyond his own behavior (this type of position can be called *non-individualistic*) but that facts about the particular individual's behavior etc. play *no* role at all in determining what his words mean.[153]   As a result, meaning-theoretic accounts often leave little, if any, room for a gap between the content of one's beliefs and the conventional meaning associated with the sentence one uses to express them. That is to say, no sense can be made of the sort of gap that Dummett and Evans, plausibly enough, take there occasionally to be between the conventional meaning of our words and the contents of our beliefs.

It should not be surprising, then, that the further someone strays from the use conventionally associated with his terms, the harder it becomes to take the meaning-theoretic line towards the interpretation of his thoughts and utterances. We may be willing to say that Bert believes that he has *arthritis* in his thigh, but this is partly because his conception of arthritis (with one notable exception) corresponds quite closely to the conventional one. We would, however, be less willing to say that, for instance, he believed that orangutans were a type of fruit drink that has been sitting in his refrigerator for the last few days, just because he uttered the words "Orangutans are fruit drinks that have been sitting in my refrigerator for the last few days."[154] In such cases we may take the belief-theoretic approach and say that he was talking/thinking about, say, orange Tang when he used the word "orangutan." Bert's conception of arthritis is enough like ours for us to understand him as being wrong about the same disease that the doctor is right about, but the errors envisaged about 'orangutans' are *so* radical and far-reaching that it becomes difficult for us to think of him as thinking about orangutans at all. In much the same way, if one clings to the meaning-theoretic account and assigns belief contents based upon the conventional meanings of the words uttered, then one ends up with

---

[153]   The "non-individualistic/anti-individualistic" terminology is taken from Katarzyna Paprzycka's discussion of a similar taxonomy of behavioral explanations (Paprzycka forthcoming).
[154] Though, some, for instance Tyler Burge, might. (See Burge 1979, 91.)

extremely unintuitive belief assignments whenever one encounters slips of the tongue, spoonerisms, or malapropisms.[155]

Still, one thing meaning-theoretic accounts explain quite elegantly is our capacity to learn what others believe from what they say. On meaning theoretic accounts, not only are belief and thought contents tied together, but these contents are shared by the entire community. As a result, the same content can be believed by a speaker, expressed by that speaker and grasped by his interlocutor with little problem. Indeed, while some of the belief assignments they endorse are extremely implausible, meaning-theoretic accounts allow for a very powerful notion of linguistic expression and communication, a notion that belief-theoretic accounts are not able to make room for. Indeed, if meaning-theoretic accounts have any problems in this area, it is that they can't make sense of our *not* coming to know what someone believes from his sincere utterances, something that, intuitively, does happen on occasion. If one does not recognize a malapropism for what it is, it seems that one should fail to learn what the speaker believes from his utterance, but the meaning-theoretic position seems to leave no room for this sort of failure of communication.

4.3 A synthetic position

Belief-theoretic and meaning-theoretic accounts are each at their strongest when their rivals are at their weakest (accounting for communication with the belief-theoretic, capturing the speaker's point of view with the meaning-theoretic). As a result, one can easily find oneself oscillating between the two depending upon which aspect of thought or language one is focusing on at the time. Nevertheless, while the belief- and meaning-theoretic accounts may be the two most obvious ways to explain the expressive

---

[155] And, once again, loses the humor in the Archie Bunker utterances.

constraint, they clearly are not the only ones. In particular, a non-reductive account that stressed the *interdependence* of belief content and linguistic meaning could, if available, capture the virtues and avoid the faults of its two reductive rivals.

Such and intermediate position will not loose sight of the belief-theoretician's insight that a good interpretation of a person will allow us to understand how he sees the world we share. The point of interpretation is to be able see *our* world through *his* eyes. Since capturing the *speaker's* perspective on the world is the goal of the interpretation, it is natural that his self-interpretations should have a certain normative force for us. We assume that the speaker knows what he means by her words, and try to make our interpretation of her match what his self-interpretation would be if he were aware of all that we were. This requires that the we try to understand the speaker as he would understand himself, and it will be argued here that it is the synthetic position, rather than the belief-theoretic one, that is best able to do this.

Such a synthetic position one should (like the belief-theoretician) give primary importance to capturing the interpretee's beliefs and point of view, but (unlike the belief-theoretician) recognize the importance of the speaker's beliefs about the public language and his relation to it as well. If this is done, one can allow the meanings conventionally associated with the speaker's words in the public language to play a large role in determining what the speaker means, while at the same time not losing the agent's point of view.

Belief-theoretic accounts often get much of their motivation from a perceived need on the interpreter's part to capture the interpretee's point of view and explain his behavior. However, a crucial part of an agent's point of view is his belief that he means the same thing by his words as his fellows do -- that he speaks a shared language.[156] This belief is manifested in behavior such as his attitude ascriptions (such as willingness to move from

---

[156] The belief that we speak the same language as our fellows need not ever be consciously formulated by the speaker; rather, it might be part of something like what Searle refers to as "the Background." (See Searle 1983, Dreyfus 1991, and appendix F.)

"A said '*P*'" to "A said that *P*"), his deference to correction, his willingness to admit that he doesn't know whether a particular object is an *X* or not, or that some things that he judges to be *X*s in good light, etc. could turn out not to be *X*s, etc.[157]

If the speaker's use strays from the social norm that he believes himself to be in accord with, then, as when we are faced with any other conflict between a speaker's beliefs, capturing his point of view should lead us to favor whichever the speaker himself would favor if the inconsistency were made manifest to him. If the speaker is unwilling to defer to the accepted usage, then we should attribute to him the false belief that he was talking about the very things that his fellows were. If he is inclined to defer, we should treat those applications of the term that are out of line with social usage as misapplications. Since we are trying to capture the agent's self-interpretation, this latter attribution is contingent upon the deference being understood by the speaker *himself* as a response to his own misapplication of the term. For instance, we can imagine cases where we would defer purely for 'pragmatic' reasons. That is to say, we understand ourselves as having meant something different by a particular word than most of our fellows, and we change how we use the term so that we can communicate with them more easily. In such cases, capturing our self-interpretation would not involve picking out the 'standard' referents with our terms. As a result, if people really did defer only for pragmatic reasons,[158] deference behavior would have little philosophical import. Still, while there certainly are cases where we alter our usage for pragmatic reasons (such as when we intentionally 'misuse' a term to communicate more smoothly with someone we know to misuse the term in a particular fashion), this is certainly not *always* the case. People frequently defer to common usage because they understand themselves as having had false beliefs about what items actually are denoted by the term in question. If we

---

[157] These topics will be discussed further in chapter three.

[158] As writers such as Davidson and Bilgrami frequently (but implausibly) argue (for a discussion of these arguments on their part, see the following chapter).

take the speaker's self-interpretation to be authoritative in these matters, then we should accept such a stance on the speaker's part at face value.

The strong *anti*-individualistic position associated with meaning-theoretic accounts, which hold that what we mean by our terms is *always* dependent upon social usage, receives no support from this argument, since there are clearly cases where we diverge from the social usage and, when this divergence is pointed out, see ourselves as mistaken about the social usage rather than the objects in question. This explains why we can usually ascribe socially determined concepts like "arthritis" to Bert, but cannot do so to the person who claims that there is an "orangutan" in the fridge. Bert himself will tend to view himself as having a false belief about arthritis, while the orangutan man will not tend to view himself as having a mistaken belief about orangutans, he will understand himself as having a misexpressed belief about orange Tang. However, a *non*-individualistic position, which only requires that *some* of our concepts *can* be tied to social use, follows from the more synthetic position outlined above. The motivation behind the belief-theoretic accounts can actually be understood as leading to this synthetic position, not the individualistic one often associated with belief-theoretic accounts.

This synthetic position can thus accept the type of non-individualist content ascriptions Davidson rejects while holding on to his insight that while "many philosophers [i.e., meaning-theorists] have tied the meaning of a speaker's utterance to what others mean by the same words," "this tie is neither essential nor direct," but rather "comes into play only when the speaker intends to be interpreted as (certain) others would be," so that when this intention is absent "the correct understanding of a speaker is unaffected by usage beyond the intended reach of his voice."[159] As a result of this, it is not always helpful to discuss issues in this area in terms of 'individualism' at all, since

---

[159] Davidson 1992, 261.

the 'non-individualistic' position outlined above is still *methodologically* individualistic.[160]

The resulting position is classified comfortably neither as belief- nor as meaning-theoretic, but it seems to capture the advantages of both types of view. On the one hand, it allows us to capture the agent's point of view just as well as, indeed better than, belief-theoretic accounts. On the other hand, it provides a good account of how we come to know what another believes from their sincere utterances. While the synthetic account provides room for a distinction between what a person means by his words and the conventional meanings associated with those words, it also explains why these will usually be the same. We generally intend to be interpreted as our fellows are. Such an intention is, of course, defeasible (if, for instance, we make a slip of the tongue or our usage strays too far from the public norm), but minor differences between ours and the public use will not be enough to make us give it up. While we occasionally may not assent to the attribution of standard meanings to our words, these cases tend to be exceptional, and we usually consent to be interpreted (and, crucially, interpret ourselves) in accordance with the conventional meanings. Indeed, the fact that Bert is speaking to a *doctor* and not, say, his brother may be relevant here. When we are actually speaking to experts, the presumption, on the part of both participants, that we will intend to be following 'expert usage' will be considerably stronger than when we are talking within a possibly idiosyncratic sub-community such as our family.[161] While the utterer and hearer may have different beliefs associated with the terms used in a particular utterance, both will usually be willing to defer to the conventional usage. As a result, both can be taken to mean the same thing by the sentence, and to have thoughts with the corresponding content. Deference to the social norm is the default case, and, as a result, we are entitled to assume that others mean just what we do by our terms unless they give us reason to think otherwise.

---

[160] For a discussion of this, see appendix D.
[161] This is discussed in greater detail in chapter three.

The expressive constraint (once again, the requirement that when a speaker expresses a belief by sincerely uttering a sentence, the utterance and the belief have the same content) can thus be incorporated into a synthetic account of the relation between thought contents, utterance contents, and the contents conventionally associated with our words. Furthermore, such an account avoids the pitfalls associated with the belief- and meaning-theoretic accounts. Both belief- and meaning-theoretic accounts work with the assumption that for conventional meaning to be even *relevant* to what our own words mean, it must be *equated* with it. Once this assumption is given up, one can give a much more plausible account of the expressive constraint. Since the tendency to understand the constraint in belief- or meaning-theoretic terms is part of the reason it can seem so hard to find an equilibrium for our semantic terms that incorporates it, this should be an encouraging result.

5. Conclusion

We discussed in the previous chapter the need to find an equilibrium for our semantic terms. This chapter has argued that the expressive constraint, by insisting that speakers have beliefs with the same content as their sincere utterances, effectively requires that whatever equilibrium we find for thought or utterance content will be an equilibrium for both. Given that this 'internal' connection between thought and utterance content makes finding an equilibrium for our semantic terms considerably harder, one might try to do without the expressive constraint. However, attempts to explain away the constraint as an overstatement of a merely *empirical* generalization about the frequent correlation between thought and utterance content were discussed and found to have serious problems. It was argued that most philosophical accounts of thought and language are committed to the expressive constraint, and that if the constraint is given up, a number of counterintuitive consequences follow. In particular, rejecting the constraint makes the intelligibility of successful communication and belief attribution problematic. Indeed, it

was argued that, if the constraint is given up, there would be little reason to think that even our *self*-ascriptions would usually be true. This last problem leads to the result that, once the expressive constraint is given up, there should be nothing wrong with the sincere utterance of a 'Moore sentence' such as "Carburetors are expensive, but I don't believe it," since one could have good reason to think that one lacked beliefs with the content of even one's sincere utterances. If one accepts the constraint, one should provide some explanation of it, and it has been argued that the two most obvious explanations of the constraint (giving an independent account of linguistic meaning and identifying thought content with it, and giving an independent account of thought content and understanding linguistic meaning in terms of it) are both inadequate. Rather, it was suggested that one can give a less reductive and more synthetic account of the constraint by stressing the interdependence of belief content and linguistic meaning that comes from the fact that our language is itself one of the things about which we have many beliefs. The importance of such attitudes about language will be stressed further in the following chapter when we discuss how the 'socially infected' ascriptions Burge appeals to can be incorporated into an equilibrium for our semantic terms.

Chapter three: Individualism, Non-Individualism, and Anti-Individualism

While it is generally accepted that the ascriptional practices Kripke appeals to (especially those involving proper names discussed in chapter one) cannot be excluded from any acceptable equilibrium, the status of those ascriptional practices Burge appeals to is considerably less well entrenched.[162] The purpose of this chapter will be to argue that our practice of interpreting people in terms of the 'accepted' usage in their community can, indeed should, be incorporated into equilibrium for our semantic terms.

One set of terminological points that should be made clear from the start is the following. The term "individualism" will be here associated with the belief that an individual whose usage is known to stray from the social norm should *never* be interpreted according to the usage of his community; "anti-individualism" will be associated with the position *contrary* to individualism, i.e., the belief that such a speaker should *always* be interpreted in terms of the accepted usage in his community, and "non-individualism" will be associated with the position *contradictory* to individualism, i.e., the belief that an individual whose usage strays from the social norm should at least *sometimes* be interpreted in terms of the accepted usage in his community.[163] It will be argued below that incorporating the practices Burge appeals to requires only that one adopt a *non-individualistic* position; it does not require adopting the stronger *anti-individualistic* position.[164]

---

[162] The practices Putnam appeals to occupy something of an intermediate position, and will not be discussed directly here. Most who are willing to accept Burge's intuitions will also be willing to accept Putnam's, and any argument that supported the practices Burge and Wilson draw to our attention (which the following two chapters will attempt to do) will also constitute an argument in favor of the Putnam practices.

[163] This terminology is, once again, taken from Katarzyna Paprzycka's discussion of a similar taxonomy of behavioral explanations. (Paprzycka forthcoming.)

[164] Burge himself is best understood as defending a non-individualistic rather than an anti-individualistic position, and critics of Burge such as Davidson and Bilgrami are also, ultimately, best understood as defending non-individualistic rather than individualistic positions (though they are clearly committed to *methodological* individualism in way a in which Burge is not). Still, while Davidson's and Bilgrami's considered positions are not actually individualistic, they can still be considered "defenders of individualism" since they do, in fact, go to considerable length to defend it (at least in part because they take Burge to be defending an anti-individualistic position). For a discussion of these issues, see appendix D.

There are, of course, other criteria by which a position could be characterized as "individualistic," and the position defended in this chapter, while non-individualistic in the terms described above (hereafter "ascriptional" individualism"), is "individualistic" in certain other important respects. In particular, while this chapter argues against ascriptional individualism, it does endorse "methodological" individualism. Methodological individualism requires only that, when interpreting an individual speaker, we count 'external' factors (such as, say, the usage of his surrounding community) as relevant in determining what he means only if the interpretee himself takes them to be so. Such a position is, in many respects, at least as interesting as ascriptional individualism, but when we are concerned with whether the actual ascriptional practices that Burge focuses upon can be accommodated into an equilibrium, ascriptional individualism is what is at issue. This is, of course, because it is only ascriptional individualism that is necessarily incompatible with the acceptability of such ascriptional practices. (Methodological and ascriptional individualism are often run together, and they are discussed and distinguished in greater detail in the appendix D.)

This chapter will first investigate which of individualism, non-individualism and anti-individualism is best supported by our ascriptional practices, and then consider various arguments for taking these ascriptional practices more or less seriously. It will then turn to our general beliefs to see if and how they can be accommodated by the three positions. It will turn out that the non-individualistic alternative is not only best supported by our ascriptional practices, but it is also compatible with most of the general beliefs frequently taken to be incompatible with it.

1 The Ascriptional Practices

1.1 The relation between the Burge and the Kripke practices

As stated before, Kripke's intuitions about proper name reference in the cases described in chapter one are almost universally shared. Even defenders of description

theories tend to accept the intuitions -- they just try to argue that they can be explained within the descriptive framework. On the other hand, the intuitions that Burge appeals to are considerably less secure, and the first reaction of many readers (including myself) when presented with Burge's thought experiments was simply to deny that, say, Bert meant what most other people did by "arthritis."

Kripke's work on proper names can thus be seen as having the potential to shore up a more generalized externalism in two ways. First of all, it can 'prime' our intuitions in a way that makes the conclusion Burge draws from his thought experiments seem more appealing.[165]   Secondly, and much more significantly, if it turned out that the best account of the phenomena Kripke points to entailed that Burge's conclusions also followed, then Burge's conclusions would share the intuitively secure footing of Kripke's work.   This chapter will argue that this is just what happens: the best explanation of the Kripke ascriptions allows one to endorse the Burge ascriptions as well.

Whether or not an equilibrium for our semantic terms would incorporate the types of ascriptions Burge appeals to[166] may depend, then, to a considerable extent upon whether or not these ascriptions are best understood as different aspects of the very same practices that Kripke appeals to. The anti-/non-individualists will thus have a considerable interest in showing that the ascriptional practices appealed to in the Burge cases are really of a piece with those appealed to in the Kripke cases. The individualist, on the other hand, will be concerned with describing the ascriptions that Kripke and Burge appeal to as instances of two distinct practices rather than as different aspects of the same practice. If he succeeds in doing this, the individualist can say that while we may have to accept

---

[165]   As I said in chapter one, I doubt that the intuitions behind Burge's thought experiments would carry the force they do if the general beliefs they cut against hadn't already been shaken by Putnam's and Kripke's work.   *Naming and Necessity* could, in some sense, be understood as being at least part of a paradigm shift in philosophical thought about language.   After this shift, appeals to the sort of thought experiment Burge relies upon had already been entrenched in the profession, and had they not already been so, the comparative infirmity of the intuitions Burge appeals to might have been enough to dismiss them.

[166]   Which, we must remember, are not the same as our reactions to thought experiments.   (See chapter 1, sect. 8.)

Kripke's deeply entrenched set of ascriptional practices, Burge appeals to a different set that can be left out of any equilibrium.

While engaging in such a project, the individualist could appeal to in the widespread assumption that, since the works of Kripke and Burge highlight different things about the contents of our words and thoughts,[167] there should be two (possibly complementary, possibly rival) 'externalist' theories of content: 'causal/historical' theories and 'social context' theories.[168]

The anti-/non-individualist would obviously like to show that the line of thought above is mistaken, and that whatever differences there are between Kripke and Burg, they can be understood as relatinglargely to the types of examples focused on, not to the general practices they draw upon. The fact that they highlight different aspects of the *same* practice has been obscured, the anti-/non-individualist will argue, by a number of factors. The first of these is that much of the excitement first generated by Kripke's work related to its prospects for providing a physicalistically respectable account of reference and rehabilitating certain types of essentialism.[169] Both of these projects seem quite alien to anything that would come out of Burge's work, and the focus on them caused the social, historical, and genetic aspects of Kripke's account to be ignored. Secondly, Burge himself initially presented the phenomena in "Individualism and the Mental" as being

---

[167] Namely, Kripke's work (along with Putnam's) illustrates that content is in part determined by our *physical environment* while Burge makes a structurally similar point relating to our *social context*.

[168] For a recent argument for the distinctness of Burge's and Kripke's positions, see Donnellan 1990. This assumption, though not often explicitly stated, is often implicit in criticism and expositions of Kripke, Putnam and Burge. Criticisms of Kripke and Putnam based on the assumption that they rely on metaphysically natural kinds presuppose that they provide a different sort of theory from Burge, who clearly requires no such assumption. Further, expositions of Burge according to which he argues that a society's present application of a term determines its meaning, and thus cannot account for Putnam's intuitions that we meant something different by "water" than our Twin-Earth counterparts even *before* water's chemical structure was discovered (Pettit & McDowell 1986b, 7-10), also presuppose that two different types of external contribution are being pointed to. Of course, this is not to deny that everyone recognizes the two points are of the same general *type* and that all the writers involved are externalists.

[169] See, for instance, Field 1972, Devitt 1980, and Salmon 1982. Putnam's work (especially Putnam 1975a), inspired similar hopes, though it seems quite clear that Putnam does not share them (see, for instance, Putnam 1981c, 1981d, 1988).

unaccountable for by the picture presented by Kripke,[170] thus making his work appear to present a rival paradigm rather than developing an earlier one. One of the main reasons, then, for thinking that Kripke and Burge are talking about different practices comes from the tendency to view Kripke as offering a plan for a non-intentional reduction of reference according to which causal relations and the 'metaphysical essences' of various kinds determine what we mean. Whatever Burge was up to, it certainly wasn't that.[171]

As a result, the anti-/non-individualist will be concerned with showing, then, that Kripke did not, in fact, present a theory in which causation and the world's essential structure could alone determine the meanings of our words.[172] The anti/non-individualist will thus stress that Kripke is not only notoriously cagey about referring to his work as a "theory" at all,[173] but he also explicitly states that his account does not eliminate reference from the explanans.[174] Kripke is perfectly willing to call his 'picture' a

---

[170] Or, for that matter, Putnam. While Burge admits that there are points of contact between his picture and Putnam's (Burge 1979, 73, 117, 118), he still believes that the argument extracted from Putnam's paper "is narrower in scope," and "seems to work only for natural kind terms and close relatives" (Burge 1979, 118). He claims that purely causal accounts cannot account for the examples he discusses for the following reason.

> The problem is unaffected by the suggestion that we specify input and output in terms of causal relations to particular objects or stuffs in the subject's physical environment. Such specifications may be thought to help with some examples based on indexicality or psychological success verbs, and perhaps in certain arguments concerning natural kind terms (though even in these cases I think that one will be forced to appeal to intentional language). . . . But this sort of suggestion has no easy application to our argument. For the relevant causal relations between the subject and the physical environment to which his terms apply -- where such relations are non-intentionally specified -- were among the elements held constant while the subject's beliefs and thoughts varied. (Burge 1979, 107.)

These claims are, in some sense, unobjectionable, but Burge would be very much mistaken if he saw them as applying to either Kripke or Putnam (rather than to their purely causal/reductive reconstruals). Recently Burge has characterized Kripke's work as showing that reference depends on "a mix of causal and intentional elements and includes a person's reliance on others to fix the referent" (Burge 1992, 24), and he seems to have mellowed a bit on Putnam as well (though this may be because he sees Putnam as having changed his views (Burge 1992, 47)).

[171] Another factor that made the two sets of practices seem different was the unfortunate tendency to understand Burge as defending the type of quasi-communitarian picture of linguistic norms that would seem to have more in common with the reading of Wittgenstein purportedly in Kripke 1982 than it does with Kripke's own views. (I have doubts about this reading of Kripke 1982 as well, but one can see instances of such quasi-communitarian readings of both Burge and Kripke's Wittgenstein in, for instance, Bilgrami 1992a.)

[172] Reasons for thinking that neither Kripke nor Putnam should be understood this way are discussed in appendix A.

[173] Kripke 1972, 93.

[174] Kripke 1972, 94, 97.

"historical" one, and it is only later writers who have put the emphasis on causality. Kripke's picture of reference involves two parts: the 'baptism' at which the word is first associated with an individual or kind, and the transmission through which the meaning is preserved through subsequent use.[175] While both aspects of this picture can be, and have been, given causal/non-intentional interpretations,[176] such readings misrepresent Kripke's text and ignore the possibility that Kripke's work could provide the foundation for a more general externalist theory.

Not only can Kripke's account of baptism be generalized beyond proper names and natural kind terms,[177] but his account of transmission can also be so extended. Kripke argues that a speaker refers to whatever he names "by virtue of his membership in a community which passed the name on from link to link,"[178] and there is nothing in this story that suggests that this linking is only possible with proper names and natural kind terms. All that is required of such a linking is that "the speaker intends to use the name in the same way as it was transmitted to him."[179] Indeed, there is little reason to think we would have such intentions for proper names and natural kind terms but not for terms

---

[175] "A rough statement of a theory might be the following: an initial 'baptism' takes place. Here the object may be named by a description. When the name 'is passed from link to link,' the receiver of the name must, I think, intend when he learns it to use it with the same reference as the man from whom he heard it." (Kripke 1972, 96.)

[176] See Devitt 1980, x.   In the case of the baptism, a purely causal reconstrual would involve either the person or the initial sample (and its 'essential' structure) entering into causal relations with the baptizer (Devitt 1980, 27-31, 138.). However, such a reading would have trouble making sense of Kripke's claims that "there need not always be an identifiable initial baptism" (Kripke 1972, 162). Furthermore, words whose meaning is the product of a reference shift could all be understood as, in some sense, lacking an initial baptism.

[177] Once again, for a discussion of this, see appendix A.

[178] Kripke 1972, 91. This appeal to the speaker's being embedded in a social/historical environment is by no means isolated (see, for instance, Kripke 1972, 94, 95).

[179] Kripke 1972, 163. Devitt considers the reliance on these reference-preserving intentions to be a weakness of Kripke's account, and tries to give an purely causal replacement of them (Devitt 1980, 138).   Of course this appeal to reference-preserving intention may seem too intentional and voluntaristic, and it may be misleading to say that when we use a name or kind term, we use it with the intention that it refer to whoever or whatever those from we learnt it referred to by it. However, much the same result can come from the speaker's having the belief that he is using the name or kind term in the same way as it was transmitted to him, and this belief, even if not explicitly formulated, can be seen as a consequence of the speaker's belief, or even presupposition, that the term is part of a shared common language that he is speaking. For a discussion of how one can get at what Kripke wants here in a less voluntaristic fashion, see appendix F.

such as "arthritis" or "contract."[180] The commitment the speaker takes on through this intention is not the type of commitment that he could only form with proper names and kind terms, and so if such intentions and commitments can determine what our names and kind terms refer to, they should be able to do so for our other terms as well.

This is the source of the following general belief about meaning:

If the speaker intends to use a term in the same way as it was transmitted to him then the object referred to by the speaker with the term is the same as the object referred to by the transmitter with the term.

As argued above, the acceptance of this belief (along with some plausible assumptions about the speaker's intentions) leads almost immediately to the acceptance of the ascriptional practices Burge appeals to. As a result, the individualist will have some reason to want to reject such a belief, though it is hard to see how to reject it without also giving up the Kripke practices. The story that is in play about the transmission of names and kind terms already allows that a speaker can use them at least partially in virtue of his assuming that the term in question belongs to a language shared between him and the other users of the term that he has encountered.[181] Once this reference-preserving mechanism is in play, there seems to be no reason for it not to be extended to terms in the language other than proper names or natural kind terms.

For instance, if John means *arthritis* by "arthritis," and Bert picks up the term from John intending to use it with the same reference as him, Bert should mean arthritis by "arthritis" as well. The fact that Bert believes that one can get arthritis in one's thigh is no more of a problem than the fact that our belief that Thales said "all is water" may turn out to be false. Since the process involved does not require that our beliefs about the referent be true in order for us to talk about it (that Thales said "all is water" might be all a speaker (explicitly) believes about him), Bert's words and thoughts are connected to

---

[180] Especially since a given speaker or community may not be aware of whether a given term of his picks out a 'natural' kind.

[181] This assumption need not, of course, be explicitly made. It may be the sort of implicit 'assumption' that forms part of what Searle describes as the "Background" against which our intentional states are found (Searle 1983). For a fuller discussion of these issues, see appendix F.

arthritis in spite of his mistaken beliefs about where it can occur. It should go without saying that this Kripkean story will predict that, in a counterfactual situation where the community that Bert has been brought up in means something different by "arthritis," Bert will as well. It should also be noted that this application of Kripke's picture explains Burge's intuitions by making an appeal only to the account of transmission. The more problematic and controversial picture of the baptism need not be brought into play, since it is simply taken as *given* in Burge's thought experiments that the members of the surrounding communities mean what they do by their terms.

It seems, then, that the practices appealed to by Kripke are best understood as encompassing the Burge ascriptions as well. Attempts to describe them as distinct (one practice tying content purely to causal history, another purely to social context) fail to do justice to the practices themselves. They also fail to do justice to the philosophers in question, as Kripke's picture comes out as unrealistically essentialist, and Burge's as overly anti-individualistic and unrealistically communitarian.

1.2 The status of the Burge practices

In addition to challenging the connection between the Kripke and the Burge practices, the individualist can also challenge the entrenchment of the Burge practices directly. There are at least three ways to do this. The first is to argue that we don't need to incorporate such attributions into an equilibrium because we do not, in fact, make attributions this way. The second is to argue that, even though we do make such attributions, they can be 'explained away' pragmatically and do not reflect our considered judgments. The third is to argue that even if such ascriptions do reflect our pre-theoretical judgments, such judgments are of little theoretical significance. It will be argued below that all three approaches are ultimately unconvincing.

1.21 On the denial of the practice

If we did not make the kind of attributions Burge claims we do, there would clearly be no pressure to incorporate such attributions into any equilibrium, since they would turn out not to be entrenched at all. One writer who has frequently argued that our ascriptional practices themselves actually favor individualistic content attributions is Davidson. He has several reasons for thinking this. The first of these is that there are cases where we clearly do seem to make content attributions individualistically, and it would certainly strengthen the individualist's position if he could show that our ascriptional practices do not uniformly support anti-/non-individualism the way they are often represented as doing. Davidson focuses on what he takes to be such cases quite frequently.

> If you see a ketch sailing by and your companion says, 'Look at the handsome yawl,' you may be faced with a problem of interpretation. One natural possibility is that your friend has mistaken a ketch for a yawl, and has formed a false belief. But if his vision is good and his line of sight favorable it is even more plausible that he does not use the word 'yawl' quite as you do, and he has made no mistake at all about the position of the jigger on the passing yacht. We do this sort of off the cuff interpretation all the time, deciding in favor of reinterpretation of words in order to preserve a reasonable theory of belief. As philosophers we are peculiarly tolerant of systematic malapropism, and practiced at interpreting the result. The process is that of constructing a viable theory of belief and meaning from sentences held true.[182]

Davidson's choice here is not a particularly happy one. With respect to the ketch/yawl example, one can freely admit that the speaker "does not use the word 'yawl' quite as you do," while still insisting that the speaker has made a (non-perceptual) mistake and does mean by "yawl" what the interpreter does. Indeed, the question of whether difference in 'use' in this restricted sense[183] must involve a corresponding difference of meaning is precisely what is at issue between the individualist and the non-individualist. Matters are not helped by the fact that the case itself is underdescribed. If the speaker is, like myself, generally ignorant of sailing terminology and sailboat construction, I would be inclined think that he simply cannot tell ketches and yawls apart and has falsely called a ketch a

---

[182] Davidson 1974b, 196.
[183] That is, in the sense corresponding to how the word is actually applied by the speaker, as opposed to the more 'generous' notion of "use" in which we can be said to use the word "arthritis" to talk about arthritis even if we occasionally misapply the term.

yawl. On the other hand, if the speaker is quite familiar with the two types of boat and *systematically* calls ketches "yawls" and yawls "ketches" (and perhaps is able to explain that the two types of boats are distinguished by the position of the jigger etc.), then the claim that he cannot tell the two types of boat apart seems clearly false and the claim that he means *ketch* by "yawl" seems considerably more plausible. If non-individualism is true and what sorts of ascriptions we make must be decided on a case-by-case basis, it may frequently turn out that a disputed example is simply underdescribed.

1.211 Slips, malapropisms and misunderstandings

Nevertheless, even though I do not share Davidson's intuitions in this *particular* case, Davidson's invocation of malapropisms should make clear the types of case he has in mind. In particular, there seem to be three types of case where we tend not to understand the interpretee, and the interpretee does not understand himself, in terms of the meanings conventionally associated with his words:[184]

(1) Slips of the tongue, slips of the pen, spoonerisms and typos.

All of these 'performance errors' are pretty clear cases where we do not interpret someone according to the meanings conventionally associated with his words. If, after I have looked at Bill's diary, I type "I was just looking at Bill's dairy,"[185] this is not taken to be an expression of a false belief that I was looking at Bill's *dairy*, but rather as a clumsy expression of a true belief that I was looking at Bill's *diary*. "Dairy" is here either interpreted as meaning *diary*, or it is interpreted as meaning dairy, but no judgement is attributed corresponding to the assertion made.[186]

---

[184] Burge discusses these three types of case as instances where we do tend to reinterpret the speaker. (Burge 1979, 90.)

[185] The sort of mistake that, as any reader of earlier drafts of this knows, I make all too frequently.

[186] In terms of the discussion in chapter two, the expressive constraint does not hold for such performance errors.

(2) Malapropisms, systematic mispronunciations and spelling mistakes

Malapropisms such as "we need a few laughs to break up the monogamy" are also cases where we are not inclined to interpret a speaker in terms of the meanings conventionally associated with his words. For instance, until quite recently, I often confused "eczema" and "emphysema." As a result, I would say things such as "Peter's smoking has made his eczema much worse." In such cases, I was not understood, and did not understand myself, as making the false claim that Peter had a skin problem that was exacerbated by his smoking. I was simply understood as meaning *emphysema* by "eczema."[187] As with slips of the tongue and mispronunciations, we are inclined to say "what he meant to say was …" rather than understanding the speaker as having a false belief.[188]

(3) Radical misunderstandings.

If his misunderstanding is radical enough, there is also a tendency to reinterpret the speaker so that he is not actually speaking and thinking falsely about the objects that his words are conventionally associated with. For instance, if, Bert came to believe that "arthritis" was a retirement community in southern Florida, we would probably be disinclined to treat his "arthritis" utterances as being about *arthritis*. The more radical the speaker's misunderstanding is, the greater percentage of his beliefs will have to be given up in order to accommodate the beliefs tying his use to communal standards. As a result, the more radical the speaker's misunderstanding, the more likely it will be that the equilibrium for his total belief set will not include these more socially directed beliefs.

---

[187] Idioms provide many cases like this. People new to a language frequently make minor but often amusing errors with their idioms. I know of a woman who once said at work "I came in late and the boss ate me out good"; she is understood by everyone as truly saying and believing that she was severely criticized verbally ("chewed out"), not as having a false belief about her sex life.

[188] We are more likely to say "by 'epitaph' I meant 'epithet'" or "by 'orangutan' I meant 'orange Tang'" than we are to say "by 'arthritis' I meant 'something very much like arthritis that one can also get in one's thigh.'"

1.212 Individualistic vs. individualized interpretations

It is, of course, no coincidence that the sorts of cases where a more individualistic interpretation seems tempting are precisely those that caused trouble for the 'meaning-theoretic' accounts of the expressive constraint discussed in chapter two. However, while such cases may force us to give up purely meaning-theoretic accounts of belief, they do not necessarily mean we should give up non-individualism or even anti-individualism. While such cases may show that the interpretation may need to be *individualized*, it need not follow that the interpretation idiosyncratic speaker must be *individualistic* in the sense of being *independent* of standard usage. (Indeed, given obvious differences of pronunciation etc. our interpretations of others will have to be individualized to some extent at at least that level.)

One could admit that the speaker's idiolect remains primary when it comes to deciding which terms are assigned to which concepts, and still claim that a speaker's concepts are determined by the usage of the community. We each speak our own idiolect, and while these idiolects may be different for each person (and each person at each time), they remain dependent upon the shared linguistic norms. Malapropisms and spoonerisms provide fairly clear examples of this. When Mrs. Malaprop says (while pointing to an arrangement of epithets) "What a nice derangement of epitaphs" one interprets her as meaning *epithet* by "epitaph." This interpretation works only for Mrs. Malaprop and, in this sense, there is an idiolect that only she speaks. Nevertheless, the account does not treat her usage as unconstrained by social norms. On such a theory "epitaph" is true not of just those things Mrs. Malaprop would call "epitaphs," but rather of those things that are, in fact, *epithets* -- where what counts as an epithet is tied to the usage of the community. One cannot just read the meanings conventionally associated with the terms in the language into the speaker's utterances, but even when we come up with a theory uniquely suited to the speaker, such a theory may still make use of socially

determined concepts.[189] Cases of the first two types (slips and malapropisms) will usually be of this nature.

Still, while the first two types of case (slips and malapropisms) may ultimately not be best understood in completely individualistic terms, it seems quite possible cases of the third sort (radical misunderstanding) may favor a more individualistic understanding. A speaker may understand a term in an extremely idiosyncratic way that cannot be mapped on to any socially determined concepts. In such cases, it may be best to interpret the speaker as using a concept that is not in the social stock. Though even here, it isn't clear that the speaker is being understood individualistically. In particular, it may be that the speaker has a novel concept partially in virtue of combining other concepts in an idiosyncratic way, but the concepts so combined may themselves be socially determined. I may be the only person to mean *beer from Ireland* by "porter," but what I mean by "beer" and by "Ireland" (and thus what I ultimately mean by "porter") may still be determined by social use.

Nevertheless, while the fact that there may be *some* cases that must be understood this way causes serious problems for anti-individualism, it does not really support individualism, since individualism requires that *all* cases be of this sort. As a result, the fact that we may, at times, interpret others 'individualistically' suggests that our ascriptional practices favor non-individualism over anti-individualism, but not that they favor individualism over non-individualism.

---

[189] Indeed, the point could be seen as an extension of Evans' discussion of linguistic "counters" below:

> When we wish to establish what a person is saying in using certain words, we must get clear exactly which linguistic counter, so to speak, the speaker is putting forward. Here the speaker's intentions are indeed paramount. But once it is clear which linguistic counter he is putting forward, the content of what he says is determined by the significance which that counter has in the game. (Evans 1982, 68-9.)

> Not only can we identify the counter the speaker intends to play in spite of the fact that he may have various mistaken ideas about the counter's value (arthritis-cases), but we can also identify the counter he has in mind in spite of his using a word that is conventionally associated with a different counter (malapropisms).

1.213 Davidson on "arthritis"

Davidson, however, claims that our ascriptional practices not only favor non-individualism over anti-individualism, but also favor individualism over non-individualism. He is thus frequently[190] concerned with arguing that we do, in fact, always interpret each other individualistically. As a result of this, Davidson suggests that when Burge claims that we would say that Bert means *arthritis* by "arthritis," he is describing his own theory-driven intuitions rather than our deeply entrenched ascriptional practices.

Davidson claims not to be impressed by Burge's arguments because he has "a general distrust of thought experiments that pretend to reveal what we would say under conditions that in fact never arise."[191] However, it is hardly fair of Davidson to argue that Burge's arguments rely on our intuitions about "what we would say under conditions that in fact never arise." While such a charge could perhaps justly be made against arguments relying heavily on, say, Twin-Earth cases, examples such as Bert's use of "arthritis" are of a sort that we run into frequently in everyday practice. I was recently with a group engaged in a conversation when it became clear that one speaker thought that "vests" could have sleeves. The speaker had a number of true beliefs about vests (that they were items of clothing, that I was wearing one at the time, that they form an essential part of three-piece suits etc.) and could recognize most vests when he saw them, but would also be willing to apply the term to 'sleeved vests' and other "jackets that weren't quite up to snuff." Nobody (including the speaker himself) took the speaker to mean anything different than usual by "vest." He was simply understood as having a false belief about vests. Similar cases come up all the time in everyday discourse (John thinks that any coffee with steamed milk is a cappuccino, Susan thinks that sherbet is a

---

[190] But by no means always (see appendix D).

[191] Davidson 1991b, 197. It should be noted in this context that Davidson's discussion of the "Swampman" is not a thought experiment of this sort. He is not trying to bolster a theory by appealing to our intuitions in this case, but using a case to illustrate the consequences of a theory he has already defended. Indeed, the example is effective precisely because what he takes to be the correct thing to say about the Swampman is *counterintuitive*.

type of ice cream, etc.) and there is no reason to think that the purported phenomena rest on intuitions about various types of 'science-fictionish' thought experiments of the sort familiar from, say, debates about personal identity.

However, Davidson does not merely argue that the types of cases Burge appeals to "never arise." He also doubts that someone "unspoiled by philosophy" would make ascriptions the way Burge says they would.[192] Davidson argues that, in precisely the type of case Burge appeals to, it is simply "false that our intuitions speak strongly in favor of understanding and interpreting an agent's speech and thoughts in terms of what others would mean by the same words."[193] Rather, he claims that our actual practice favors the individualistic interpretation. He gives an example of and motivation for such 'correct' interpretive practices in this passage:

> Suppose that I, who think the word 'arthritis' applies to inflammation of the joints only if caused by calcium deposits, and my friend Arthur, who knows better, both sincerely utter to Smith the words "Carl has arthritis"… If Smith (unspoiled by philosophy) reports to still another party … that Arthur and I both have said, and believe, that Carl has arthritis, he may actively mislead *his* hearer. If this danger were to arise, Smith, alert to the facts … would add something like "but Davidson thinks arthritis must be caused by calcium deposits." The need to make this addition I take to show that the simple attribution was not quite right; there was a relevant difference in the thoughts Arthur and I expressed when we said "Carl has arthritis."[194]

According to Davidson, then, our ascriptional practices, even in the sorts of cases Burge focuses on, will actually be individualistic.

However, with respect to the claim that such unqualified reports could mislead a speaker, it is, of course, open to the anti-/non-individualist to respond that their misleadingness is not incompatible with their truth. Such reports truly convey the thought expressed, but can give a misleading impression of the subjects *total* state of mind. If I tell someone who needs to go downtown that I have a car, I will have misled him if I do not also tell him that it is broken. The fact that such unqualified reports are misleading can be attributed to the pragmatics of the conversational situation rather than the content of what is said. Given that I know that the person wants to go downtown, the

---

[192] Davidson 1987, 449.
[193] Davidson 1991b, 197.
[194] Davidson 1987, 449.

speaker will assume that I would not have mentioned the car if it was non-functional. If we are told that someone has a car, we will assume that it is in working order unless told otherwise. In much the same way, since most speakers know that arthritis is a disease of the joints, a hearer will assume that the person with an arthritis-belief will have such knowledge unless he is informed otherwise. As a result, one can mislead such a speaker by making the attribution without the qualification.

The extent to which such qualifications are necessary may be part of what determines the desirability of ascribing a socially determined concept. For instance, if Davidson believes that arthritis is caused by calcium deposits, but that is his only false belief about arthritis, then most of his behavior relating to arthritis (his purchase of arthritis medicine, his belief that he has it in his knee, his expression of sympathy to a friend who claims to have it, etc.) can all be explained with straight attributions, and additional qualifications about his beliefs about its cause will not be necessary. Of course, some of his arthritis behavior might need the added qualification (his cutting down his milk consumption because be believes he is getting arthritis), but this will probably be only a small fraction of it. On the other hand, for someone who believes that an orangutan is a fruit drink, it will turn out that just about *all* of his 'orangutan' behavior (his claiming that their is a cold pitcher of it in the fridge, his desire for some when he is thirsty, his claim that they can be purchased for $1.25 at the corner store, etc.) will require that we mention the idiosyncratic belief, so the attribution of the 'standard' concept is considerably less useful.

Davidson is aware, of course, of such appeals to the pragmatics of the conversational situation, and he claims that they overlook "the extent to which the contents of one belief necessarily depend on the contents of others" and forgets that "thoughts are not independent atoms."[195] However, there seems to be no reason to think that the non-individualist (rather than the anti-individualist) needs to deny this. Such an appeal to

---

[195] Davidson 1987, 449.

pragmatics does not, as Davidson suggests, commit one to there being a "simple, rigid, rule for the correct attribution of a single thought."[196] Unlike both the individualist and the anti-individualist, the non-individualist can take things on a case-by-case basis, so the claim that there should be no such rigid rule could be understood as favoring non-individualism.

Davidson here seems to be assuming that holism entails that any difference of belief will lead to a difference of meaning, and if he were right about this, holism probably would entail individualism.[197] There is, however, little reason to believe this.[198] If what a term refers to is determined by the equilibrium for all of the beliefs and applications associated with it, a change in a single belief will frequently make no difference as to what the equilibrium will be. One can thus admit that what a speaker means by his terms is a function of all of his beliefs, without being committed to the claim that any change in his beliefs would constitute a change in what he means. This would only follow if the function in question is the identity function, and while some versions of 'conceptual role semantics' may work with such a function (and thus bring with them the counterintuitive consequences frequently associated with holism), the equilibrium driven story from chapter one clearly has a more 'normative' function in mind. As a result, while the account of meaning in chapter one is holistic, it avoids the counterintuitive consequences frequently associated with holism by both critics like Fodor and supporters such as Davidson.

Non-individualistic accounts can take best advantage of the flexibility that accompanies the radical interpretation process by dealing with utterances on a case-by-case basis and thus taking to heart Davidsonian claims such as "there can be no simple, rigid, rule for the correct attribution of a single thought,"[199] and "when others think

---

[196] Davidson 1987, 449.
[197] This is, of course, an assumption he shares with many of holism's harshest critics. (See, for instance, Fodor and Lepore 1992)
[198] See the discussion of holism in chapter one for a fuller discussion of what comes immediately below.
[199] Davidson 1987, 449.

differently from us, no general principle … can force us to decide that our difference lies in our beliefs rather than our concepts."[200] While a Davidsonian position certainly could not be anti-individualistic, it need not be individualistic either. A non-individualistic position seems in keeping with both the spirit and much of the practice found in Davidson's writings.[201]

1.22 'Explaining away' the practice

Thus far we have been looking at attempts to show that we do not, in fact, ascribe attitudes in the way that Burge suggests. There is, however, another line of attack that, while admitting that we do make ascriptions the way Burge claims that we do, tries to downplay the significance of this part of our practice.[202] Doing so will usually involve trying to 'explain away' phenomena that seem to give at least *prima facie* support to anti-/non-individualistic positions. Attempts to explain away such phenomena can be seen in discussions of both our belief ascriptions and our deferential practices. (The two are not unconnected, since our deferential practices are a reflection of the crucial case of our self-ascriptions.)

1.221 Belief ascription

Davidson, for instance, has argued that, even if Burge is right about what we would say about Bert (which, as noted above, Davidson doubts), such ascriptions can be understood as reflecting laziness on the interpreter's part or pragmatic constraints on 'off the cuff' interpretation rather than actual truths about what he means by his terms. Given the number of comparative strangers we meet each day, there is no real possibility of

---

[200] Davidson, 1974b, 197.

[201] Indeed, while Davidson tends to view himself as an individualist, his reasons for doing so tend to focus almost exclusively on problems with *anti*-individualistic positions, and Davidson's own position (as distinct from his explicit pronouncements on the subject) is perhaps best understood as non-individualistic. (For a fuller discussion of this, see appendix D.)

[202] "We would, in everyday reporting of belief, say "Bert believes that he has arthritis in his thigh." . . . It is quite open, however, whether this fact about everyday reports has anything to do with the issue before us. The issue before us is a philosophical one." (Bilgrami 1992a, 74.)

engaging in full fledged radical interpretation with each of them. Because of this, we save time and effort by attributing 'standard' contents to their words.[203]   As Davidson puts it: "We do not have the time, patience, or opportunity to evolve a new theory of interpretation for each speaker … But if his first words are, as we say, English, we are justified in assuming he has been exposed to linguistic conditioning similar to ours."[204] "When some of these assumptions prove false, we may alter the words we use to make the report, often in substantial ways. When nothing much hinges on it, we tend to choose the lazy way: we take someone at his word, even if this does not quite reflect some aspect of the speaker's thought or meaning."[205]

However, this explanation of our ascriptional practices as being a natural response to the lack of evidence we usually have in conversation has problems accounting for the fact that we often stick with these 'lazy' ascriptions even after learning that the interpretee uses a term differently than we do. Indeed, if we do choose to flag any differences, we tend to flag them in terms of the speaker's beliefs, not his meanings. Our tendency to interpret people non-individualistically cannot, then, be explained away by the claim that it is the product of time-constraints on off-the-cuff interpretations. Even if they had all the time in the world to interpret Bert, many interpreters (including, obviously, Tyler Burge) would still attribute to him the false belief that he has arthritis in his thigh.

Perhaps even more damaging to the claim that the non-individualistic elements in our ascriptional practices are produced by such pragmatic constraints is the fact that our *self-*

---

[203] "This is not to deny the practical, as contrasted with the theoretical, importance of social conditioning. What common conditioning ensures is that we may, up to a point, assume that the same method of interpretation that we use for others, or that others use for us, will work for a new speaker. We do not have the time, patience, or opportunity to evolve a new theory of interpretation for each speaker, and what saves us is that from the moment someone unknown to us opens his mouth, we know an enormous amount about the sort of theory that will work for him -- or we know we know no such theory. But if his first words are, as we say, English, we are justified in assuming he has been exposed to linguistic conditioning similar to ours (we may even guess or know the differences). To buy a pipe, order a meal, or direct a taxi driver, we can go on this assumption. Until proven wrong; at which point we can revise our theory of what he means on the spur of the moment. . . . Knowledge of the conventions of language is thus a practical crutch to interpretation, a crutch which, under optimum conditions for communication, we can in the end throw away, and could in theory have done without from the start. " (Davidson 1982, 278-9.)

[204] Davidson 1982, 278-9.

[205] Davidson 1987, 449.

ascriptions are often made along non-individualist lines. It is very hard to reconcile the claim that non-individualistic content ascriptions are based upon our not taking the time to find out just what the interpretee really believes with the apparent fact that, in spite of our knowing what we believe, we make such non-individualistic ascriptions to ourselves. We may know perfectly well what we believe, and we still frequently see ourselves as having *misapplied* words and as having had *false* beliefs when we find that our usage deviates from the social norm. If we really self-ascribe contents non-individualistically, then the attempt to explain away non-individualistic content ascriptions by constraints on off-the-cuff interpretations seems broken-backed. The appeals to our 'everyday ascriptions' and deference behavior should thus be understood as part of a mutually supporting argument, with deference being crucially important not for its own sake, but rather as an indication of our self-ascriptions.[206]

It should not be surprising, then, that defenders of individualism have often tried to argue that we do not, in fact, make such self-ascriptions. To do so, they must give an account of what happens when we defer to standard usage that does not involve our understanding ourselves as having had false beliefs or as having misapplied our own terms. Such accounts are the topic of the next section.

1.222 Deference

Defenders of individualism will argue that deference is explained not by the fact that speakers understand themselves as having misapplied their terms, but rather by the fact that speakers know that, if they wish to be understood, they should use their words as others do. This allows deference to be given a pragmatic explanation.[207] These more pragmatic and historical explanations of deference also open up an account of how we can be *mistaken* when our usage strays from the norm. We have a general intention to

---

[206] Deference is thus connected to a crucial sub-case of our content attributions, and should not be understood as part of a separate, and perhaps unrelated, argument against individualism, as Bilgrami suggests Bilgrami 1992a, 78.

[207] See, for instance, Bilgrami 1992a, 79-89, Davidson 1994a, 9.

speak as others do, and so we can be considered to have made a mistake each time we fail to realize this intention.[208] These mistakes, however, consist in our failure to communicate effectively, not in our failure to understand how our words should be used.[209]

The gloss that individualists give to linguistic norms makes them much weaker than those proposed by Burge. In particular, they constitute a set of *hypothetical* rather than *categorical* imperatives for the speaker. That is to say, all of the 'semantic norms' in these cases are contingent upon the speaker having particular *desires* (usually having to do with communicating effectively). These desires do not, in themselves contribute to the term's meaning. Bilgrami characterizes deference as a response to a norm "which necessarily has a hypothetical formulation,"[210] and Davidson claims that "any obligation we owe conformity is contingent on the desire to be understood."[211]

However, the picture of the norms involved as being hypothetical in this fashion seems to misrepresent the phenomena we are dealing with. While there are undoubtedly cases where we alter our usage in order to promote ease of communication, such cases should not serve as a general model of what is happening when we 'correct' our usage.[212] As a general psychological claim, saying that deference is a response to such a hypothetical norm relating to our desire to communicate is not especially plausible. When our usage is corrected, and we accept the correction, we usually understand ourselves as having made a mistake about the items in question ("I didn't know that you could have oral contracts" etc.). We do not usually understand ourselves as having meant something different by the word, and as changing our usage just because the new usage is more communicatively efficient. People could come to see their mistakes this way

---

[208] Bilgrami 1992, 110.

[209] See Davidson 1994a, 9, Davidson 1986a, 434, Bilgrami 1992a, 110.

[210] Bilgrami 1992a, 110. He goes on immediately to specify the norm as saying "speak like others do, *if it pays to do so*," or "more specifically," "I ought to use words as others do, *if I want to be easily understood*." See also Bilgrami 1992a, 111 for the claim that the norms involved are not only hypothetical, but also extrinsic (i.e.: make no references to, say, 'English').

[211] Davidson 1994, 9.

[212] Such a denial of generalizability is also found in Burge 1989, 184.

(Davidson and Bilgrami apparently do), but most speakers do not. When students get wrong answers on vocabulary tests, they are understood, both by their teachers and by their peers, as not knowing what the word in question means, not as speaking a different language that is poorly suited for doing well on tests. This understanding of our usage is not just a product of, and certainly extends beyond, the classroom.[213]

The extent to which Davidson overinstrumentalizes our thoughts about language can be seen more clearly when we consider that Davidson portrays even small children as following pragmatic strategies when they learn their language: "The child babbles, and when it produces a sound like 'table' in the evident presence of a table, it is rewarded; the process is repeated and presently the child says 'table' in the presence of tables."[214] Needless to say, there is no reason to think that the child needs any kind of reward system to encourage it to speak as others do in its environment. A child does not imitate the way those around her speak because she *wants* to be understood, or because she wants to speak, say, English properly. Imitative behavior begins at the phonetic level almost as soon as babbling begins, and speaking as others do seems to be inherently desirable for the language learner and is not dependent upon any other desires.[215] Indeed, the categorical nature of the imperative is such that it is almost misleading to say that the child *wants* to speak as others do at all; no such general desire needs to be present, each instance of imitative behavior is desirable in itself, not by reference to any more general desire. Even though it is the behavior of adults that we are interested in, this fact about infants illustrates, in a particularly pure form, a tendency to defer to surrounding usage that carries on throughout our lives.[216] This is not, of course, to deny that our actual need

---

[213] Which is not to say that our education plays no role in how this understanding manifests itself.

[214] Davidson 1992, 263.

[215] Of course things are a little more complicated than this, since issues of identification play some role in this process, but the type of sociolinguistic issues having to do with language learning and social and gender identification will not be explored here.

[216] Furthermore, even if one were to admit that the imperative to bring one's use into line only presents itself as a hypothetical one, the norm of speaking, say, English correctly seems just as likely to be psychologically active as a norm related to ease in communication that makes no reference to a shared language. This becomes especially clear when we consider those who consciously stray from a type of usage that may now facilitate communication because of their belief that various new idioms do not

to communicate plays an important role in our understanding of deference behavior. However, while pragmatic factors may be involved in explaining deference behavior, allowing such factors to play some role is not incompatible with being a non-individualist. Pragmatic factors may shape which attitudes we have without these factors themselves being explicitly represented by us. Language would not be much use if we did not speak in roughly the same way, but these more 'evolutionary' considerations need not manifest themselves in any actual *desire* to speak as others so that we can be understood. Other sets of beliefs and desires can produce the same results just as, if not more, effectively.[217]

For instance, one could explain defference by an appeal to a sub-personal mechanism that causes us discomfort whenever our behavior does not conform with that of those around us. We need not have any concious *desire* to conform with our peers, we just find ourselves doing so. Our concsious beliefs and desires may just rationalize, rather than cause, the conformist behavior in question (in this case, defference to social usage). One could imagine the evolutionary use of such a mechanism, but it should be stressed that this does not count as a 'pragmatic' explanation of defference behavior of the sort that the individualist needs. If the pragmatic factors are not at the concious level, and the agents actual beliefs and desires are still in terms of the agents having made a factual mistake about the objects in question rather than communal usage, then both the conditions of application and general beliefs associated with our words will favor the non-individualistic equilibrium.

It would, of course, be possible for a community to get together and simply decide to try to speak the same way, with certain members given the right to determine the

---

constitute part of "proper English." Indeed, the widespread belief that languages are being 'corrupted' testifies to the presence of a conception of language which plays a normative role independently of its role in communication. This belief in the possibility of linguistic corruption may just be a reflection of our pre-theoretical conception of language, but, right or wrong, it is this pretheoretical conception that determines the motivational nature of the linguistic imperatives acting upon us.

[217] Just as a set of agents with a number of deontological commitments may be better at, say, maximizing the greatest happiness for the greatest number, than a group that worked explicitly with such a maxim.

standards that the other members will follow, but the same result is often produced more effectively by the shared belief that there is a language which all of the members are speaking, and a right or wrong way of speaking it. The attitude most people have towards the apparently arbitrary norms of their culture and time (styles of dress etc.) makes it clear that we are born with a tendency to think that we are not merely conforming to some arbitrary standard, but rather that there are 'objective standards' that we are collectively tracking. Our conception of language is the product of this objectifying response when exposed to linguistic behavior. This conception forms a background against which deference takes place, and seems, most of the time, entirely natural. It is a background in which there are such things such as the French language and the English language, and there are such things as speaking these languages correctly and incorrectly. These assumptions are frowned upon by linguists, often for good reason, but they play a role in the unreflective behavior of even those speakers who have been taught that these background assumptions are not justified. The concept of meaning is tied up in a conceptual web with this pretheoretical conception of language, and the associated conception of meanings as being generally shared by all speakers of a language -- which is responsible for our "slovenly" attributions of content -- is a result of this.[218]

Deference is thus important not because shared concepts are somehow needed to provide a causal explanation of our deference behavior;[219] rather, a community's having

---

[218] Issues relating to this are discussed, to a certain extent, in appendix F.

[219] Something like this understanding of the appeal to deference is suggested by Bilgrami in the following:

> The argument from deference proposes that we explain it by adopting Burge's social externalism about concepts and contents. The claim is that the fact that Bert (and all of us) so readily changes his ways to conform to the standards of his community and its experts is to be explained by saying that he all along had the expert's concept. He was just misapplying it. He now has come to have a fuller understanding of it. (Bilgrami 1992a, 78-9.)

While Burge does point out deference as an important phenomenon, it is difficult to find in "Individualism and the Mental" anything like the claim that deference is a phenomenon that we need shared concepts to explain. A *causal* account of deference behavior could still be provided in individualistic terms. The 'argument from deference' is not that Bert's changing his ways "to conform to the standards of his community and its experts is to be explained by saying that he all along had the expert's concept" (Bilgrami 1992a, 78-9), but rather that Bert's changing his ways to conform to the

the right sort of deference is partially constitutive[220] of their having a shared system of concepts. The difference between the 'right' and the 'wrong' sort of deference can be understood in terms of speakers' self-interpretations, and whether or not the imperative to defer is perceived as being hypothetical or categorical.[221]

The attempts to 'explain away' deference and our belief ascriptions through appeals to pragmatic considerations seem unsuccessful, since they presuppose an indefensible picture of the interpretee's psychology. As a result, our ascriptional practices seem to favor non-individualism over both individualism and anti-individualism.

1.23 Dismissing the practice

Finally, the individualist can argue that our everyday ascriptions do not carry significant weight when discussing philosophical/theoretical issues in the theory of meaning. A fairly clear example of this attitude can be found in Bilgrami's *Belief and Meaning*:[222]

> The appeal to ordinary practice of reporting seems hardly to acknowledge that the question here is a thoroughly philosophical one. The notion of content which is the subject of philosophical question is a notion which has a theoretical status because the very issue underlying the philosophical question is a theoretical one: what notion of content to posit such that it will serve in explanations of behavior as well as stand in essential relations with the external world? Why should we take facts about ordinary, everyday reports to be relevant to this theoretical notion? If Burge's examples tell us anything, it is that ordinary, everyday practice is neglectful of the fairly serious differences between Bert and his fellows.[223]

This seems to deny that our ascriptions have even *prima facie* weight.[224] However, the mere fact that one is looking for a *theory* to account for the truth of some of our semantic terms' general characterizations does not give one the right to simply dismiss the *practice*

---

standards of his community is to be explained by his having beliefs (about doctors' knowing more about arthritis than him, etc.) that also allow us to say that he has misapplied his concept if his usage does not correspond the community standard.

[220] In Davidsonian terms, we could say that, their having a shared system of concepts is a matter of their being best interpreted as having such a shared system, and their deferring as they do (with the corresponding self-interpretation) contributes to just such an interpretation being the best.

[221] The fact that we do defer -- and see ourselves as responding to categorical rather than hypothetical imperatives -- plays a crucial part in a number of arguments against the individualist's appeal to general beliefs in section 2.

[222] Once again, it should be noted that while Bilgrami could be understood as defending individualism here, his actual position is best understood as ascriptionally non-individualistic.

[223] Bilgrami 1992a, 74. See also pp. 5-6.

[224] "Ordinary chat … is not in itself of relevance or interest in the philosophical study of meaning." (Bilgrami 1992a, 75.)

that constitutes the putative extensions of those terms. Unless, of course, one is willing to simply change the subject (which some have accused Bilgrami of doing).[225]

Still, our actual practice of using our terms is not sacrosanct. For terms like "witch" or "phlogiston" it seems as if our past applications were entirely mistaken. Even though our ascriptions suggest that *other things being equal* a non-individualistic interpretation would be better, if other things are not equal (in particular, if philosophical/theoretical considerations favor the individualist), they should not carry any decisive weight. It is possible that enough deeply entrenched beliefs are incompatible with non-individualism that it could not be accepted into any equilibrium, and this possibility is the subject of the next section.

2. The general characterizations

The fact that deeply entrenched aspects of our ascriptional practices are non-individualistic does not entail that there will be a non-individualistic equilibrium available for our semantic terms. While the fact that we make non-individualistic ascriptions provides an at least *prima facie* case for thinking that such ascriptions are correct, the individualist can still argue that, even if such ascriptions make up part of the putative extension of our semantic terms, they could not be part of any final equilibrium between our ascriptional practices and the general characterizations we associate with our semantic terms. The individualist will insist that individualism cannot be given up once we see how closely it is linked to the general beliefs about content. The debate between individualism and anti-/non-individualism has often been understood as between those motivated by a desire to stick closely to our ascriptional practices (the putative extensions) and those motivated by more 'theoretical' concerns (the general characterizations).[226] If it turns out that equally well entrenched general characterizations

---

225 See Travis 1995, 144.
226 Bilgrami, quoted above, is the most explicit about characterizing the debate this way, but the view is fairly widespread. Arguments for non-individualism often amount to little more than a statement of

are incompatible with non-individualism, one might not be able to find *any* equilibrium for our semantic terms, and some form of bifurcationism would seem necessary. There is, then, considerable reason to want to accommodate as many of the general beliefs associated with our semantic terms as possible within a non-individualist framework. This section will argue that many of the most central of these beliefs can, in fact, be so accommodated. Not only can many of the general beliefs that seem incompatible with non-individualism be understood in a way that is compatible with it, but the understandings of such beliefs that remain incompatible with non-individualism are also incompatible with the Kripke cases. Furthermore, there are a number of general beliefs that fit better with non-individualism than individualism.

2.1 We speak a shared language

One general belief that clearly seems to fit better with the non-individualist picture is the belief that we speak a shared language (ie: for most speakers we meet, we feel that we are speaking the same language as them -- though some may speak it differently). While such a belief is clearly quite deeply held, it is not at all obvious what it is. Just what is it to share a language in the first place? The individualist, with his commitment to each speaker having his own independent idiolect,[227] cannot take language to be shared in any robust sense. As a result individualists tend to come up with an *ersatz* notion of a shared language, based on sufficient similarity, or simply reject the notion of a robustly shared language as an obviously confused idea. Indeed, the two responses tend to go together. Individualists claim that we understand ourselves as speaking a shared language in a way that is obviously false, and the *ersatz* notion is all that can be salvaged from the linguistic phenomena.

Burge's thought experiments, and those against usually depend upon its incompatibility with various deeply entrenched beliefs we have about self-knowledge, behavioral explanation etc.

[227] Once again, the important part of his position relates not to our each having our own idiolect, but to the *idependence* of these idiolects.

The individualist will argue that while people tend to believe that (almost) everyone in the community speaks the same way, they greatly underestimate the differences between speakers (differences that can be, after all, considerable).[228] In light of this, the individualist will recommend that we understand 'sharing' a language in a way that does not require that the languages involved be *exactly* the same (since facts about pronunciation etc. (not to mention malapropisms etc.) make it clear that no two people speak in the same way). Rather, a 'language' should be understood as 'shared' by a population if the idiolects of the speakers in the population are similar enough for them to communicate effectively, they learn how to speak from one another, are 'corrected' by each other, etc. Languages would thus be shared to a greater or lesser degree depending upon how much the idiolects overlap, with a limiting case of a language being fully shared by groups of speakers whose idiolects are indistinguishable. This limit would, of course, probably never be reached, so there would remain a sense in which languages are never fully shared.[229]

 Nevertheless, even those people who are willing to recognize differences in pronunciation often tend to think that we at least *mean* the same thing by our words as our fellows, and the assumption that we refer to the same things by our words as our fellows is not as obviously false as the corresponding idea about pronunciation.

This may be to a large extent because, as the discussion of malapropisms above should have made clear, what a speaker means by his terms can depend upon the usage in his community even if the actual (phonetic) word he attaches to the concept is different form the one so attached in his community. If I mean *epithet* when I say "epitaph" (and thus the meaning of "epitaph" in my idiolect is determined by how "epithet" is used by

---

[228]   Linguists often can find informants who will insist that everyone in the community pronounces the words in the language the way they do, and when the linguist provides alternate pronunciations found among other informants, the informant in question will insist that such forms are incorrect and that no one speaks that way.

[229]   "If we are precise about what constitutes a language, it is probably the case that no two people actlully do speak the same language." (Davidson 1992, 260, see also Chomsky 1988 36-7, Lewis 1969, 201.)

my community), there is a clear sense in which I share a language with my fellow speakers. Even if we are not speaking the same way, there is still a "division of linguistic labor" in such cases. This "division of labor" metaphor suggests a way in which speakers can share a language without speaking the same way.[230]

Indeed, not only is speaking the same way not necessary for sharing a language, it is not sufficient either. Imagine, for instance, that two historically isolated communities lived in the same environment, but never interacted with each other (and were, perhaps, even unaware of each other's existence). By sheer coincidence, the two communities came to speak in exactly the same way, applying the same words to the same things etc. Now, even if one were to admit that these speakers spoke *exactly* the same language,[231] there would remain a sense in which they did not *share* a language at all. One can understand languages as shared in the sense that I 'share' a haircut with everyone who has a haircut sufficiently similar to mine (a sense in which I probably 'fully' share my haircut with no one), but one can also understand languages as shared in the sense that a cooperative is shared by its members. The individual member has a stake in the cooperative, and the actions of the other members determine just what cooperative he has a share in in a way in which other people who 'share' my haircut do not affect the haircut I have. It is in this latter, more robust, conception of sharing that the two communities do not share a language at all, while members of the same community who speak differently can share a language in this sense.

In spite of its 'robustness' this conception of sharing a language is, if anything, more attainable than the individualist's similarity-based notion according to which no two

---

[230] The division of linguistic labor is something the ascriptional (rather than the methodological) individualist cannot really account for.

[231] And this admission is not one we should clearly make. One could compare the discussion in chapter one of the disagreement between a molecular and an evolutionary biologist over what counts as being of the 'same species.' Something molecularly identical to our dogs but having no evolutionary connection to them would count as a dog for the first but not for the second. In much the same way, a language that had the same phonology, syntax and even semantics as English, but which had a completely different origin, might count as being the same language as English to a formal linguist but a different language to a historical one.

people 'fully' share a language. I 'share' a language with other members of my community because I look to their help in deciding how to apply my own terms and they do the same with me. Because of this, once the idea of sharing a language is tied to our shared practice and less closely to the idea of speaking in exactly the same way, the idea that we share a language seems neither obviously false, nor accountable for in individualistic terms. We can understand people having the same language in terms of their sharing a language, or understand people sharing a language in terms of their speaking in the same way. The former strategy clearly seems to be the more workable.

While the non-individualist can claim that we really do *share* our language, since we rely upon others to specify the referents of our terms etc., the individualist's understanding of the shared language belief requires that we understand it in a way that is obviously false (i.e.: we all speak in exactly the same way) and then tone it down so that it is roughly true (we all speak in roughly the same way). The individualistic understanding of the shared language belief requires, then, that we understand it as false. Since the belief is comparatively well entrenched, this should count against individualism.

2.2 Our terms have determinate extensions

Another general belief that may not sit well with individualism relates to the assumption that the extensions of our terms are more determinate than our discriminating capacities: we sometimes do not know whether a term applies to a certain object or not. This belief is phrased in metalinguistic terms, but it is manifested in non-metalinguistic attitudes such as, say, realizing that, while one is sure that a particular bird one is looking at (in good light etc.) is a raven or a crow, one does not know which.[232] Indeed, I initially wasn't even sure if there was a difference between these two types of birds. I asked

---

[232] See also appendix F for a further discussion of how such attitudes could be manifested.

around, and while eveyone was sure thaere that was a difference, no one was sure what it was (actually got conflicting suggestions about which one was bigger).

Such a belief causes problems for the individualist because there are many cases where the speaker will not be willing to say that a given term definitely does or does not apply to a given item. These may occasionaly be colloquial way of reporting vaguness ("I'm not sure if Bill is bald or not"), but frequently they are not. If content is specified individualistically, and I am inclined to say of a particular bird, "I don't know whether that bird is a crow or a raven," then my term "crow" should be neither true nor false of that particular bird. If we determine content individualistically, many of our terms will be undefined for items we intuitively take to be either in or out of the term's extension.

This fact is obscured if we focus on cases like Bert, who has a fairly specific, albeit idiosyncratic, conception of what arthritis is. Individualism would seem more plausible if we believed that people were more semantically decisive than they often are. The fact that people's discriminatory capacities seem to underdetermine what we intuitively take them to be referring to may, ultimately, be a more serious problem for the individualist than the fact that they often misidentify what is intuitively taken to be the referent of their terms. Not only will the individualist leave us with thoughts of an intolerably general character, but he will have to treat us as simply confused when we claim not to know whether, say, a particular bird is a crow or a raven. Such claims presuppose that what a crow is extends beyond our dispositions to classify things as crows or not, and the individualist must say that, if I am not disposed to apply the term or its negation to the bird in question, I am mistaken in thinking that there is a fact of the matter as to whether or not a particular bird is a crow.

There may, of course, be cases where our uncertainty about whether to apply the term corresponds to an uncertainty about whether the term applies at all. We may be unsure whether, say, a habit is a type of dress, but we may be inclined to think that this kind of uncertainty corresponds to an indeterminacy or vagueness of the facts in question. Such

cases, seem, however, quite distinct from my not knowing whether a particular bird is a crow or an raven even though I'm quite sure that its one or the other. The former case is one where we may think that, even for someone who has fully mastered the term, whether it applies to a given item is not settled. The latter case is one where we are inclined to think that we have just not fully mastered the term. The individualist would have to assimilate all uncertainty to the former type, while the two types of case seem quite distinct.

This belief that the extensions of our terms are more determinate than our discriminating capacities should not be confused with (though it can lead to) the belief that our terms have completely determinate extensions. This stronger belief requires that every concept either does, or does not, apply to any possible object.[233] This stronger claim has seemed to many to express a belief that should be just given up.[234] It is

---

[233] The following examples from Frege and Derrida are fairly representative of how widespread this train of thought can be in philosophical thought about language:

> This involves the requirement as regards concepts, that, for any argument, they shall have a truth-value as their value; it shall be determinate, for any object, whether it falls under the concept or not. In other words: as regards concepts we have a requirement of sharp delimitation; if this were not satisfied it would be impossible to set forth logical laws about them. (Frege 1891, 148.)

> What philosopher ever since there were philosophers, what logician ever since there were logicians, what theoretician ever renounced this axiom: in the order of concepts (for we are speaking of concepts and not of the colors of clouds or the taste of certain chewing gums), when a distinction cannot be rigorous or precise, it is not a distinction at all. (Derrida 1977, 123-4.)

[234] It has been a target of Wittgenstein's (especially in the discussion of "familly resemblances" in Wittgenstein 1953.), and Searle (1977), among others, has characterized it as simply a hangover from positivism. Furthermore, prototype-based accounts of our concepts (see Lakoff 1987, Rosch 1975) are based on precisely the assumption that the Frege/Derrida hypothesis is false and that our concepts do not have such determinate extensions. For instance, Jackendoff stresses that there is considerable evidence that our application of words is often governed not by our having a set of necessary and sufficient conditions for a term's application, but rather by our having a prototype associated with the word, with the result that we are comfortable applying the word to items that match the prototype closely and are increasingly unsure about the term's application the further items get from the prototype in question. (Jackendoff 1988, 1990.) While we certainly need not follow Jackendoff in abandoning the extensional approach to semantics, even within the extensional approach, the prototype phenomena could be seen as affecting the shape of the resulting theory. In effect, the prototype theory might suggest that most terms be treated as vague or 'fuzzy,' and that whatever extensional analysis one comes up with for terms like 'bald' or 'heap' should be extended to the other terms in the speaker's idiolect as well. The resulting theory would seem to predict and explain his behavior perfectly well. Nevertheless, Frege would be the last person to be impressed by the purported findings of prototype theories, and he might, with some justification, argue that psychological facts about how we are disposed to apply our terms no more determine what concepts are than psychological facts about how we are disposed to make inferences determine which inferences are, in fact, good ones. All

important, however, to recognize that the weaker claim, that our terms' extensions are more determinate than what our discriminatory capacities can pick out, is enough to cause problems for individualism; one does not need the stronger claim that all of our terms have completely determinate extensions. One need not commit oneself to the claim that we have no vague terms in the language to think that, even if I only know that gold is "an expensive yellow metal," I mean more by "gold" than "expensive yellow metal."

2.3 There is a fact about what we mean

One problem with tying what we mean to social usage is that it may seem to undermine the factuality of meaning. If what we meant were determined purely by facts about our dispositions or composition, then the facts about what we meant would be tied exclusively to things about which there seem to be settled facts. On the other hand, if meaning is tied to what our community means, then if there is no fact about, say, what community we belong to (what 'sociolect' we are speaking), then there will be no fact about what we mean.

This problem is heightened by the fact that anti-/non-individualistic positions are often associated with advocating 'the primacy of the sociolect.' Such accounts suggest that the norms that the individual's idiolect is responsible to are determined by 'the usage of the community,' 'accepted usage,' 'the public language,' 'the sociolect,' etc. If the speaker's idiolect represents his "always partial, and often in part incorrect, understanding of his language,"[235] then his 'language' should be something that can be specified independently of his idiolect. Such accounts suggest that we should be able to give an account of just what these public languages are, and doing so is notoriously difficult. The non-individualistic picture suggests that there will be a single 'community' whose 'use' determines the sociolect that all of its members speak. But such 'languages,' while they are supported by our idea of, say, French and English being different

---

these facts could show is that our linguistic behavior seriously underdetermines which language we speak and which set of concepts we are, in fact, using.

[235] Dummett 1975, 31.

languages, are notoriously hard to individuate. Davidson points out that "there is a problem of deciding what group is to determine the norms,"[236] and Chomsky elaborates on a similar point:

> In ordinary usage . . . when we speak of a language, we have in mind some kind of social phenomenon, a shared property of a community. What kind of community? There is no clear answer to this question. We speak of Chinese as a language, whereas Spanish, Catalan, Portuguese, Italian and other Romance languages are different languages. But the so-called dialects of Chinese are as varied as the Romance Languages. We call Dutch a language and German a different language, but the variety of German spoken near the Dutch border can be understood by speakers of Dutch who live nearby, though not by speakers of German in more remote areas. The term "language" as used in ordinary discourse involves obscure sociopolitical and normative factors. It is doubtful that we can give a coherent account of how the term is actually used. [237]

Not only are public languages hard to individuate, but the idea that each of us belongs to a group of people all of whom speak the same language is not as easy to justify as it might at first seem. The people I defer to may not, in fact, be the same people I recognize as speaking the same language as me. The idea that there may be a single self-standing community all and only the members of which speak the same language may have to be given up. Who an individual defers to may be different for any given term, and whether the relevant 'experts' will always be in the same community is doubtful. The experts I would ultimately defer to for the use of my scientific terms may be members of the international scientific community who do not even speak English, while the people I am willing to defer to for the use of other terms may not extend beyond my city or even my block.[238] There may be no 'community' the use of which can be taken as authoritative for all my terms, and there may be no other person who recognizes just the same experts for all his terms as I do. It might seem, then, that there can be a separate practice associated with each of our words, and that a given speaker can be a part of the same practice as his neighbor with respect to some of his words while not with others.

---

[236] Davidson 1991b, 197.
[237] Chomsky 1988, 36-7.
[238] Once again, Chomsky presents this worry in a particularly succinct fashion: "Suppose the expert happens to speak a variety of Cockney that I can barely understand? Is that part of my sociolect? I can perfectly well defer to my Italian gardener with whom I share only the Latin names for elms and beeches. Is his Naples dialect part of my sociolect?" (Chomsky, letter to Bilgrami, cited in Bilgrami 1992a, 266.)

This does not mean, however, that there are any serious problems for non-individualistic conceptions of content. Rather, it just shows that such conceptions of content should be divorced from a naive picture of what a 'sociolect' must be, or even from the idea that the 'sociolect' is an essential part of understanding the 'social' aspect of language. There may be a fact of the matter as to which group a speaker would ultimately defer to for each of his terms without there being a recognizable group to which the speaker belongs all of whom speak exactly the same 'sociolect.' Non-individualism and the 'social character of language' require only that we at least occasionally defer to, and hold ourselves responsible to, the usage of others. It does not require that those others be understandable as speaking the same 'sociolect' as we do.

Of course it may also be the case that there may be no fact of the matter as to who the speaker intends to defer to, and in such cases, the anti/non-individualistic picture would seem to suggest that there is no fact of the matter as to what we mean. However, in such cases, the claim that there is no fact of the matter as to what the speaker means has some plausibility. For instance, if a speaker is unaware of the fact that doctors and lawyers define "narcotic" differently and is equally inclined to defer to each group of experts, there may be no fact of the matter as to what the speaker means by the term, since the two expert communities define the term differently. Such cases, where a term is defined differently by different expert communities, need not be that uncommon. One could also cite physicists vs. chemists on what counts as a "molecule," nutritionists vs. botanists on what counts as a vegetable, British and American English on what counts as an "endive," etc.

2.4 The explanation of behavior

Perhaps one of the most deeply entrenched general characterizations related to the belief that our behavior is explained by, among other things, our beliefs and desires.[239] While one may occasionally be able to deal with a conflict between a term's putative extension and general characterization by treating the relevant aspect of the general characterization as simply mistaken (just as, say, the belief that fish were a natural kind could be simply given up), it would be hard to take this line with the claim that our behavior is explained by our beliefs and desires. Most philosophers (the perhaps not most psychologists) who urge rejecting *this* belief would probably deny that there are beliefs and desires at all.[240] As a result, if our non-individualistic ascriptional practices really were incompatible with our using belief ascriptions to explain our behavior, then we would have to either give up those practices, accept some sort of bifurcationism, or deny that we have any beliefs at all. Behavioral explanation, then, might provide the most serious obstacle to finding a non-individualistic equilibrium. The belief that our attitudes explain our behavior is extremely well entrenched, but it is also often taken to be incompatible with non-individualism.

One line of argument to the effect that non-individualistic ascriptions cannot be used to explain behavior runs as follows: (1) whatever (causally) explains our behavior must supervene on our internal states, and (2) if non-individualistic belief ascriptions were correct, then beliefs would not supervene upon our internal states, so (3) if non-individualistic ascriptions were correct, beliefs would not (causally) explain our behavior.[241] However, just because mental states are not individuated by their causal

---

[239] This should not, of course, be understood as an empirical generalization of the type of explanation we actually give when we explain someone's behavior. Indeed, many, if not most, such explanations may make no explicit reference to beliefs or desires. "He's hungry," "She's drunk," "It's her day off," "His father just died," "She's a vegetarian" etc. are all perfectly good explanations of behavior even though they make no explicit reference to an agent's beliefs or desires. Rather, the claim is that our beliefs and desires at least *can* serve as elements in a (causal) explanation of our behavior.

[240] For instance, some 'eliminative materialists' advocate denying that there really are beliefs and desires precisely because they take 'belief/desire theory' not to explain our behavior nearly as well as its hoped-for cognitive-psychological successors. (See Churchland 1981, Stich 1983.)

[241] Fodor, for instance, argues raises just such considerations in the following passage:

powers, it does not follow that they cannot figure in causal explanations. Chairs are not, after all, individuated by their causal powers, and this does not prevent there being causal explanations involving chairs. Just because 'relational facts' determine what counts as a "chair" it doesn't follow that there cannot be causal explanations involving such terms. After all, it is a perfectly good causal explanation to say that John died because he was run over by a Mercedes Benz, even if being a Mercedez Benz does not supervene on the physical properties of the car that hit him.

Another reason for thinking that individuating content non-individualistically provides belief contents poorly suited for the purposes of behavior explanation, a reason that does not presuppose a particular picture of causal explanation, run roughly as follows. If the contents of an agent's thoughts are not tied to his conception of what he purports to be thinking about, and his behavior is explained by his conception of things, then those thought contents will not explain his behavior. Bert's behavior, for instance, might seem better explained by attributing to him beliefs involving *tharthritis* rather than *arthritis*, because *tharthritis* more closely matches his conception of what "arthritis" is. To take another example from Burge, if an agent has a conception of what a "contract" is that involves the belief that contracts must be in writing, then attributing to such an agent a desire to have a *contract* with his neighbor will not explain his insistence that he and his neighbor write down their agreement, though attributing to him a desire for a *thontract* (just like a contract but must be written) will.

These considerations are, however, far from convincing. First of all, the examples presented above might suggest that there will always be a set of things corresponding to the agent's conception (the set of all aches in the joints and limbs for *tharthritis*, and the set of all written contracts for *thontract*). However, this is not always the case. Not only

---

If you are interested in causal explanation, it would be mad to distinguish between Oscar's *mental* states and Oscar2's; their mental states have identical causal powers. But common sense deploys a taxonomy that *does* distinguish between the mental states of Oscar and Oscar2. So the commonsense taxonomy won't do for the purposes of psychology. Q.E.D. (Fodor 1987, 34.)

is the agent's conception often too vague to pick out any determinate set of objects (see 2.2), but there is often *nothing* that corresponds to such a conception. For instance, in addition to believing that other people have beliefs about "contracts," the "contract" believer may also believe that, say, there are classes in contract law taught at law schools, etc. Since there are no classes taught in *thontract* law, there will, in fact, be nothing that corresponds to his "contract" conception. If any set of items is going to be picked out for his term, part of his conception will have to be understood as mistaken, and the non-individualist will argue that the belief that contracts must be written is precisely the belief that should be given up. So the non-individualist can argue that his attributions can actually do a *better* job at capturing the agent's conception of things than the individualist's.

Furthermore, the purported problems with behavioral explanation are already present once we accept the ascriptional practices that Kripke appealed to,[242] and the problem that there may, in fact, be nothing that corresponds to the agent's conception is most evident in the case of proper names. For instance, if I believe (falsely) that Bert is a doctor, my calling Bert if I break a finger might seem to be better explained by a belief about *Thert* (i.e.: someone who fits my conception of Bert), but that still doesn't change the fact that I have a false belief about *Bert* and not a true belief about some 'intentional entity' who happens to fit my conception of Bert.[243] Whatever problems non-individualistic assignments produce for behavioral explanation are already there when we assign truth-conditions in accordance with non-negotiable intuitions about proper-name reference.

Nevertheless, the fact that the problems with behavioral explanation are also caused by accepting Kripke's work on proper names does not show that they are not real problems. And if the purported problems could not be dealt with, then the non-negotiable Kripke-intuitions and the equally non-negotiable belief that our behavior can be

---

[242] Much the same could be said of the Putnam practices (our 'water' behavior being purportedly better explained by attributing a concept that picks out both H2O and XYZ).

[243] Indeed, even if there were someone else who 'matched' my conception of Bert, I would still be referring to Bert.

explained by our beliefs and desires would force us at least to split up belief content. As discussed in chapter one, this would involve arguing that there are, in effect, two types of content: a "narrow content," that is individuated internally (individualistically), and a "wide content," that is individuated externally (non-individualistically).[244] The former is meant to capture the agent's perspective on things and thus explains his behavior; the latter captures our perspective on what the agent says and thus corresponds to what we evaluate as true or false.[245]

Since we are trying to find a single equilibrium, this is clearly not the kind of solution we are looking for. However, while such a position clearly isn't an ideal one, some have argued that it couldn't even be a *possibly* acceptable. Bilgrami has criticized such solutions for sacrificing what he calls the "unity of content," namely, the idea that it is the very same contents that are externally determined and evaluated as true and false that serve in the explanation of behavior.[246] And his 'unity thesis' can be understood as an insistence that some single equilibrium be found, if any are to be found at all. If the unity thesis really is non-negotiable, then the prospect of finding *separate* equilibria for capturing truth conditions and point of view seems fairly grim.[247]

Given that this is the case, it may be worth taking a closer look at whether non-individualistic content attributions really don't allow us to explain behavior. In particular, one might hope that there is a way to capture the agent's conception of things while still individuating his concepts non-individualistically. There may be a way to *reconcile* non-individualism and behavioral explanation, without *isolating* them from

---

[244] Various forms of which are found in, for instance, Putnam 1981a, Block 1986, McGinn 1982, Loar 1985, and Field 1978.

[245] Such accounts presupose, of course, a particularly restrictive account of how the behavior to be explained can be characterized. In particular, narrow contents can serve to explain what an agent is *trying* to do, but they seem poorly suited to explain mistaken behavior or what an agent succeeds in doing (correctly calling a tree an "elm", etc.). (For a discussion of this, see Brandom 1994.)

[246] "What makes it possible for intentional states to have a role in the commonsense psychological explanation of be understood as criticizing the idea that there could be a separate equilibrium capturing our 'point of view' which was isolated from concerns relating to what this point of view was a point of view on.

each other the way the bifurcationist does.[248] Perhaps the most obvious way to do this is to make a more extensive appeal to the agent's beliefs. While the above-mentioned speaker's insistence that he get his neighbor to sign a contract in writing may be explainable by attributing to him a desire for a *thontract*, the behavior seems to be explained equally well by attributing to him *both* the desire for a contract *and* the belief that contract must be in writing. My phoning Bert when I have a broken finger is explained not by a desire to call *Thert*, but rather by a desire to phone Bert and a belief that Bert is a doctor. Non-individualistically individuated beliefs can still explain behavior as long as they are taken together with other beliefs.[249]

Bilgrami, however, claims that such accounts, like those that rely on a distinction between narrow and wide content, are bifurcationist and thus violate his unity constraint.[250] However, while such accounts do separate the roles of meaning and belief (the first giving truth conditions and the second explaining behavior), it does not follow that Bilgrami's criticism is justified. While the unity thesis seems to match at least our pre-theoretical intuitions about the role of belief content in behavioral explanation,[251] it only requires that the externally determined contents be the "*same* contents which go into the commonsense psychological explanation or rationalization of behavior."[252] This requirement has both a weak and a strong reading. The weak reading is that the contents

---

[248] Reconciliation would show how non-individualistically specified content can be used to explain behavior, while isolation would rely on dividing the explanatory work types of content do in such a fashion that non-individualistically specified contents need play no role in behavioral explanations.

[249] Something like this position seems to be taken by, say, Davidson and McDowell. Davidson's actual position on this issue is not entirely clear and on questions of Davidson exegesis, I will be following Bilgrami at this point (see Bilgrami 1992a, 177-80 for a discussion of remarks made by Davidson in conversations relating to this issue). Furthermore, while Davidson uses this strategy to reconcile behavioral explanation with non-social externalism, he does not use it to defend social externalism, which, of course, he doesn't endorse.

[250] "[For McDowell], there are *concepts* and *conceptions* obtained from looking at the truth theories and beliefs respectively. How can one deny that this amounts to bifurcationism, with one notion of content containing the concept and the other, the conceptions?" (Bilgrami 1992a, 172.)

[251] Though Stich suggests (in the last chapter of Stich 1983) that it may turn out to be false, and that the 'beliefs' that are part of our verbal reports are not in the same system as those that explain our behavior. If this turned out to be true, and one did not, like Stich, think that this meant that there were no beliefs, one would have a type of explanatory bifurcationism in which the contents of both components were externally determined.

[252] Bilgrami 1992a, 3.

that are externally determined serve as *part* of an explanation of the agent's behavior. The strong reading is that the externally determined theory of meaning for an agent should provide *all* the resources needed to explain the agent's behavior: if you know what he means by a word, you know all his associated beliefs[253] as well, and thus can explain all his behavior.[254] It should be clear that those who propose dividing up the explanatory work between narrow and wide content violate the unity thesis in both its weak and its strong forms, and are thus bifurcationists in both senses. However, while someone who splits the explanatory work between belief and meaning may violate the strong version of the unity thesis (since he maintains that a theory of belief is also needed to explain behavior) he need not violate it in its weaker form. It *is* the externally determined contents that go into the explanation of behavior; they just do so by providing the content of the beliefs in the belief component. The conceptions associated with the belief theory are in no sense independent of the contents associated with the truth theory, and thus there is not the type of independence associated with theories that distinguish narrow and wide content. So the question becomes, then, which form of the unity thesis should we accept? Given that both the strong and the weak versions of the unity thesis are able to account for the intuition that "what makes it possible for intentional states to have a role in commonsense psychological explanation of behavior … is the very fact of their possessing content which is externally constituted,"[255] the weaker version seems preferable. While the goal of making a person intelligible is a generally admirable constraint on a theory of interpretation, it is not clear that one should want to apply it directly to the theory of meaning. After all, it is quite unintuitive to claim that simply knowing what an agent means by his terms is enough to understand, and possibly even

---

[253] Or perhaps all the beliefs relevant to a particular 'explanatory locality.' (Bilgrami 1992a.)

[254] Though, of course, the thesis is often (but not in Bilgrami's case) accompanied by the corollary that we never 'really' know what anyone means by their words and that 'complete' communication is impossible.

[255] Bilgrami 1992a, 15.

predict, his behavior.[256] It is much more natural to think that we can predict or explain an agent's behavior because we have some idea of what he believes, and this would correspond to our having a theory of belief rather than a theory of meaning. Of course, by constructing the type of theory of meaning that is in line with the strong unity thesis, one would also come to know all the agent's beliefs as well. So the issue is not whether we need to know the agent's beliefs, but rather whether or not these beliefs should be kept theoretically distinct (but not isolated) from the meaning theory.

This brings us to a crucial difference between theories that divide the explanatory work between narrow and wide content and those that divide it between meaning and belief: while the former posit a split within single folk semantic/psychological categories such as meaning and content, the latter mirror the commonsense conception that we mean certain things by our words and we have certain beliefs about the world and that while these two phenomena are related, they are distinct.[257] As a result, while strongly unified theories have an advantage over bifurcationist ones in that they can give a unified account of pre-theoretically unified subject matter like meaning, they have a disadvantage compared to weakly unified accounts in that they also unify subject matters that are pre-theoretically distinct.[258]

However, it may seem that pointing out that we can distinguish what a speaker means from what he believes may not, in itself, be enough to allow non-individualistically individuated beliefs to explain behavior adequately. Having the content of an agent's

---

[256] The idea that the content of a single belief should be enough to explain a certain piece of behavior seems undermotivated. Admittedly, single beliefs and desires are cited to explain behavior (my opening an umbrella is explained by my belief that it is about to rain etc.), but this explanation usually works against a potential background of other beliefs and desires (that I believe umbrellas stop water, that I don't want to get wet etc.) that must be presupposed for the explanation to work. We don't have any problem assuming such a background, so there seems to be no reason to want to incorporate all this background information into the content of the belief itself, especially since such an incorporation leads to a number of serious problems that will not, however, be discussed here (see my "Comments on Bilgrami's *Belief and Meaning*" MS 1993).

[257] This need not, of course, involve a commitment to any sharp analytic/synthetic distinction.

[258] Of course, theoretical progress may necessitate redrawing some of the boundaries on our pre-theoretical conceptual map, and it is always possible to claim that the dictates of theory determine how the conceptual map is to be drawn. Nevertheless, if Bilgrami is to take this line against the weak unity theorist, he will undermine his own case against the bifurcationist.

beliefs determined by things other than his conceptions would seem to allow us to attribute beliefs to him that do not seem to make sense of his behavior. For instance, if Bert really believes that he has arthritis, rather than *tharthritis*, then he believes that he has a rheumatoid ailment that affects the joints exclusively, but attributing *this* belief to him will make no sense of his insistence that there is nothing wrong with his joints. Appealing to his belief that one can get 'arthritis' in the thigh does not help if this can be understood as the belief that you can get a rheumatoid ailment exclusively affecting the joints in the thigh. The mere fact that we can appeal to other beliefs, then, does not necessarily show that there is nothing unproblematic with non-individualistic belief attributions.

This is, once again, a phenomenon we also see with proper names. If Jocasta is Oedipus's mother, and Oedipus intends to marry Jocasta, then Oedipus intends to marry his mother. Yet, since Oedipus also intends to do whatever he can to avoid committing incest, his eagerness to enter into such a marriage seems mysterious. If we are to make sense of his behavior, then, it may seem that we must deny that Oedipus intends to marry his mother, and doing this may seem to return us to attributing only beliefs that are in line with the agent's conception of the objects of his beliefs.[259] However, while many explanations of behavior require that such a constraint be satisfied, behavior is occasionally explained in a way that explicitly violates the constraint. For instance, we can explain why Oedipus married his mother by pointing out that he believed both that, since she was queen of Thebes, his mother would make a great match, and that his mother was not a relative of his (and certainly not his mother). Oedipus himself might reject such characterizations (claiming instead that he believed that, since she was queen of Thebes, *Jocasta* would make a great match, and that *Jocasta* was not a relative of his),

---

[259] Bilgrami formulates such a requirement in his "constraint C": "When fixing an externally determined concept of an agent, one must do so by looking to indexically formulated utterances of the agent which express indexical contents containing that concept and then *picking that external determinant for the concept which is in consonance with the other contents which have been fixed for the agent*." (Bilgrami 1992a, 5.)

but there is a perfectly good sense in which such ascriptions explain his behavior (which consisted, after all, in his marrying *his mother*).

Ascriptions of the sort just discussed are, of course, commonly referred to as *de re*, and are more 'perspicuously' put on the following lines: "Oedipus believed of *his mother* that *she* would make a great match" and "Oedipus believed of *his mother* that *she* was not a relative of his." Such *de re* ascriptions are clearly not constrained by the speaker's conception of things, and the ascriptions that are so constrained are typically referred to as *de dicto*.[260] Both types of ascription are used to explain behavior (the *de dicto* ascriptions explaining what he is trying to do (marrying Jocasta), the *de re* ascriptions explaining what he does (marrying his mother)), and just as we can say that Oedipus believes of his mother that she would make a great wife, we can say that Bert believes of *a rheumatoid ailment exclusively affecting the joints* that he has *it* in his thigh. If we take an agent's conception of things to constrain the type of *de dicto* ascriptions we can make of him rather than constraining the contents of the beliefs themselves, then one can individuate beliefs non-individualistically while still allowing that there are no true *de dicto* belief-ascriptions that do not match his conceptions.

It may seem that relying on the *de dicto/de re* distinction in this way also produces a type of bifurcationism that violates the unity constraint: the beliefs as ascribed *de dicto* explain behavior while those ascribed *de re* are evaluated as true or false. However, not only can we evaluate beliefs as ascribed *de dicto* as true or false, and use the beliefs we ascribe *de re* to explain behavior, but the two types of ascriptions should also not be understood as corresponding to two types of belief. If the 'bifurcationism' is *perspectival* rather than *substantive*, it isn't clear that the unity constraint is violated. Remember, the unity constraint only requires that the externally determined contents are the "*same* contents which go into the commonsense psychological explanation or rationalization of

---

[260] I.e., if the *ascribee* would accept the characterization of the belief, it is *de dicto*, if the *ascriber*, but not the *ascribee*, accepts the characterization, it is *de re*. (See Dennett 1982, Brandom 1994.) For a fuller discussion of this, see appendix E.

behavior,"[261] and the fact that these contents do fill their various roles when viewed from different perspectives doesn't change the fact that it is the same contents viewed in each case.[262]

It seems, then, that there is no compelling reason to think that non-individualistically individuated thought contents are unable to serve in behavioral explanations, and thus no reason to think that this issue should prejudice us towards believing either that there could be no equilibrium points or that any equilibrium points would have to be individualistic.

2.5 Self-knowledge

Non-individualistic content ascriptions are also often taken to be incompatible with is the deeply entrenched general belief that we are authoritative about the content of our beliefs and utterances. Non-individualistic content attributions are supposed to be incompatible with our having knowledge of the contents of our own thoughts, since they allow that the content of the our beliefs can be determined by facts of which we are unaware. If what we mean by our words and the content of our thoughts are determined by contingent factors that we may know nothing about (e.g.: what the atomic structure of gold really is, how "arthritis" is used in the community), then we often don't seem to be in a position to know what we mean. As Bilgrami puts it:

> The idea is that a pair of internally identical twins on earth and twin earth respectively have different but, as far as *they* can tell, indistinguishable substances, which they both call "water," in their environments; so, given externalism, they have different 'water'-concepts and 'water'-thoughts. Since they have different thoughts without really being able to tell the difference, they do not fully know what their own thoughts are.[263]

The claim that we know what we mean may seem easier to dismiss than the claim that our beliefs explain our behavior.[264]   The idea that we have such 'privileged access' to

---

the contents of our thoughts has occasionally been dismissed as just a relic of our Cartesian heritage that, like much of that heritage, we should rid ourselves of as quickly as possible.[265] While such a gambit might have some plausibility to someone who cannot tolerate any position that even *reminds* him of Cartesianism, it does not seem, ultimately, to be very convincing. One wants a picture in which we have self-knowledge in *some* sense, and so finding an interpretation of self-knowledge that does not involve an appeal to Cartesian privileged access is preferable to using the rejection of such access as an excuse for giving up self-knowledge.There has been no shortage of papers generated by this topic,[266] but perhaps the most prominent defender of the compatibility of non-individualism and self-knowledge is Tyler Burge. Burge argues that, since the very same factors that determine what we mean also determine what we take ourselves to mean, the fact that we may not be authoritative about these factors does not prevent us from being authoritative about what we mean.

Burge argues for their compatibility in the following fashion:

> Knowledge of one's own mental events … consists in a reflexive judgment which involves thinking a first-order thought that the judgment itself is about. The reflexive judgment simply inherits the content of the first order thought. Consider the thought, 'I hereby judge that water is a liquid.' What one needs in order to think this thought knowledgeably is to be able to think the first-order, empirical thought (that water is a liquid) and to ascribe it to oneself, simultaneously. Knowing one's thoughts no more requires separate investigation of the conditions that make the judgment possible than knowing what one perceives.[267]

Such 'homophonic' explanations of self-knowledge seem in many ways unconvincing. All such arguments tell us is that our second order judgments about our thoughts and assertions about what we mean will be both true and warrented (since what determines

---

[265] There has also been, of course, considerable evidence from psychology that suggests that we are not nearly as authoritative about our thoughts as we might have thought.

[266] See, for instance, Falvey and Owens 1994 and the papers cited therein.

[267] Burge 1988, 656. Davidson makes essentially the same point in Davidson 1989, 168, and Davidson 1994b, 235. However, given that Davidson endorses this sort of argument, it isn't clear why he thinks that there is "a conflict between Burge's social externalism, which ties a speaker's meaning to an élite usage he may not be aware of, and first person authority" (Davidson 1991b, 197), since the argument form he uses above works equally well for social externalism.[268] For a discussion of similar problems relating to attempts to explain our knowledge of language as knowledge of an "M-theory", see Dummett 1975. It should also be noted that, as Joe Camp has stressed, even if the homophonic account were able to account for our positive self-knowledge, we would still lack semantic self-knowledge in the *negative* sense. We could not be certain that any of our claims not to believe a particular proposition would be true.

the contents of these second order judgments and assertions is manifestly what determines the contents of the first order thoughts and utterances they are about). However, the truth and warrant of such second-order judgments and assertions does not amount to self-knowledge about our first-order thoughts and assertions unless we know what we are thinking and saying when we make our second-order judgments and assertions. Someone who is inclined to think that non-individualism threatens our knowledge at the first level is unlikely to admit that such knowledge of our second-level thoughts and utterances is unproblematic. The homophonic argument merely seems to push the problem back a level.[268]

That the homophonic argument may not preserve all of the self-knowledge we intuitively take ourselves to have can be seen in the fact that it still leaves open the possibility that we can't be certain that we are thinking at all. If we are thinking, our self-ascriptions will have access to the same contents as our thoughts, but if we simply seem to be thinking, then our self-ascriptions are just so much empty noise.[269] Still, this type of skeptical worry about lack of self-knowledge is not peculiar to non-individualistic positions; rather, it is the product of any 'externalistic' account.

Perhaps a more serious worry about the compatibility of non-individualism and self-knowledge that also does not seem to be addressed by the homophonic defense of self-knowledge given above is that non-individualistic ascriptions seem to require that the interpretee have thoughts that, were he both rational and aware of their content, he would not have. As a result, if one wants to hold on to the idea that the interpretee is rational, one must understand him as not being fully aware of the content of his thoughts. If Bert really believes that he has a disease of only the joints in his thigh, then either he is

---

[269] This is, of course, why the anti-skeptical arguments commonly attributed to Putnam and Davidson are unacceptable as 'straight' solutions to skepticism about our knowledge of the external world. (The attribution is considerably more plausible in Davidson's case than Putnam's.) If one is a Cartesian, one can be sure that one is thinking, but within Putnam and Davidson's framework, one has no such guarantee. (This is abundantly clear in Davidson 1987 443-44, and Putnam 1981c, 17.) The extent to which semantic externalism can be incorporated into an argument against skepticism is discussed further in my imaginatively titled "Externalism and Skepticism" (MS 1996).

irrational or he doesn't know what he is thinking. Since we have no reason to think that

he is irrational, we must assume that he doesn't know what he is thinking.[270]

However, it should be clear that we would have the same problem with ascriptions

such as "Oedipus believes that his mother is not his mother." Furthermore, it should also

be clear that, if *de dicto* and *de re* ascriptions are distinguished the way they were in the

previous section,[271] there will be no problem with non-individualistic ascriptions

resulting in our ascribing to the interpretee blatantly inconsistent *de dicto* beliefs. The

constraint on *de dicto* attributions (that they involve characterizations of the objects of

belief that the interpretee accepts) will ensure that, provided the interpretee is rational, the

beliefs attributed will not be obviously inconsistent.[272] Our 'self-knowledge' of our

beliefs as ascribed *de dicto* becomes, from this perspective, rather trivial, since it is built

into the structure of the *de dicto* ascription process that we would be able to recognize

and accept the ascribed beliefs as our own. We would not, or course, have any special

knowledge or 'first person authority' about our beliefs as they are ascribed *de re*, but this

follows directly from the account as well.[273]

---

[270] This line of thought is found most clearly in Bilgrami. See, for instance 1992b, 240.
[271] And appendix E.
[272] One can thus make ascriptions non-individualistically and still accept most of the following claim of Bilgrami's:

> What is it that you are attributing when you say that he believes he has arthritis in his thigh, what does 'arthritis' *in the de dicto specification of the content* tell us? … The prohibition of substitutions in de dicto attributions turn on the attributee *not knowing* of the co-extensiveness of the expressions involved; and I am precisely protesting that there must be some, at the minimum one, and often more, expressions that can be substituted, or else it is not clear that *any* concept of arthritis should be attributed to him. And moreover if we admit that 'a disease of the joints only' is not one of the expressions to be substituted for 'arthritis' because the attributee does not know of the co-extensiveness, then … that puts into doubt the relevance of Putnam's externalist concept of arthritis to de dicto attributions. (Bilgrami 1992b, 245-6, see also, Bilgrami 1992a, 43.)

> It is only the last sentence that the non-individualist need have problems with. If *de dicto* ascriptions corresponded to a special kind of belief, then showing that *de dicto* attributions must be constrained by the agent's 'point of view' would seem to show that there was a type of belief content that was so constrained. However, if *de dicto* ascriptions are only a type of *ascription* for the *very same* beliefs that can be ascribed *de re*, then constraints on these *de dicto* ascriptions would not entail constraints on the beliefs themselves. Because of this, while his ignorance that one can't get arthritis in one's thigh prevents one from substituting "a disease of the joints" into *de dicto* ascriptions of 'arthritis' beliefs to Bert, the substitution is allowed when we make *de re* ascriptions of those same beliefs, and thus a far less 'constrained' externalism is permitted in characterizing the contents of his thoughts.

[273] One might, of course, argue that this avoids the problem only by introducing another type of bifurcationism, but, once again, it is not at all clear whether a *perspectival* bifurcationism of this sort is

Still, there is *some* sense in which the non-individualist can allow that we remain authoritative about even our *de re* beliefs. In particular, while Bert may not know that arthritis affects only the joints, the fact that he means *arthritis* by "arthritis" is still determined by other attitudes of his -- his believing doctors know more about arthritis than he does, his willingness to understand himself as mistaken about arthritis when he is corrected, etc. If he lacked such attitudes, there would be little reason to claim that he had a made a mistake about *arthritis*. Unlike the anti-individualist (who will see Bert as mistaken no matter what his other attitudes are), the non-individualist leaves what determines whether Bert is to be interpreted in accordance with communal usage or not to be determined, ultimately, by facts about *Bert*. In this sense, Bert is still authoritative about the content of his beliefs even as specified *de re*.

Finally, it should be pointed out that ascriptively individualistic accounts of content are themselves surprisingly unfriendly to self knowledge. After all, as discussed in 1.221 and 1.222, our attributions of content to ourselves are guided by just the same principles as the attributions of content we typically make when interpreting others. Bert is more likely to say that he used to believe falsely that he had arthritis in his thigh than he is to say that he had a true belief to the effect that he had *tharthritis* in his thigh and a false belief that he meant what others did by the term. Because of this, while the individualist may avoid saying that speakers are ignorant of some aspect of the content of their beliefs, he seems forced to say that they are often (especially in cases relating to deference or the acquisition of new information) mistaken about the content of their beliefs and about what beliefs they have had in the past. If Bert did not have *any* beliefs about arthritis until he was corrected by his doctor, all of his claims about past beliefs involving 'arthritis' will turn out to be *false*. While mistakes about what one meant in the more distant past can occur, cases like Bert's being corrected would involve him being

---

at all incompatible with the unity thesis. Especially since it corresponds with a *de re/de dicto* distinction that is motivated independently of trying to reconcile non-individualism and self-knowledge.

mistaken about what he meant right up until the time of the actual correction. Indeed, the fact that he will defer (and, crucially, defer categorically) when corrected suggests that he didn't even know what he meant at the time of utterance. If we did defer for merely pragmatic reasons and did operate under hypothetical imperatives to speak as others do, then individualistic content attributions would not pose a threat to self-knowledge. It is only because deference cannot always be explained the way Davidson and Bilgrami propose (see 1.222 above) that self-knowledge becomes a problem for the individualist. Had we always deferred for the sorts of reasons that Bilgrami suggest, there wouldn't be a problem about our not knowing what we meant in such cases for the individualist, since our self-ascriptions would be correct by the individualists lights.

Indeed, there is good reason to think that individualism is ultimately *more* harmful to our conception of self-knowledge than non-individualism. On the non-individualistic account, Bert's claiming that he has arthritis in his thigh does not need to be explained by his not knowing what he meant by "arthritis" -- it can simply be explained by his having a false belief, namely, that one could get arthritis in one's thigh. It is not, however, clear that the individualist can claim that Bert's mistaken assertion that he used to mean *arthritis* (rather than *tharthritis*) by "arthritis" can be explained by anything other than his having a mistaken belief about what he meant; there seems to be no available alternative belief about the world that could be used to explain his behavior.[274]

This brings us back to something that came up in section 1.23. Bilgrami was unimpressed by appeals to our typical attributions of beliefs and desires, since he believes that when working on the *theory* of meaning, such rough-and-ready attributions should not necessarily be taken seriously.[275] However, the requirement that a theory must

---

[274] Bilgrami seems right to say that it is "foolish and unintuitive" to "grant that a wholly *non-*psychological phenomenon, i.e., an entirely abstract philosophical doctrine about reference, should be seen as raising an obstruction to self-knowledge" (Bilgrami 1992b, 236), but it may be the case that it is his own theory, and not non-individualism, that provides such an obstruction.

[275] Such an attitude is connected to Bilgrami's commitment to putting the 'theory' back into the theory of meaning, a commitment evident in passages such as the following.

attribute beliefs that are compatible with self-knowledge undermines the possibility of taking such an unashamedly theoretical stand. Our 'everyday understanding' of others will be very much like our everyday understanding of ourselves, and so any theory that takes our everyday understanding of others as irrelevant runs the risk of providing a theory according to which we systematically misunderstand ourselves, and thus lack self-knowledge. Self-knowledge thus provides a constraint that prevents the appeal to the theoretical status of the theory of meaning from being applied in too *ad hoc* a fashion.

It seems, then, that considerations of self-knowledge, while often presented as one of the stronger considerations in favor of individualism, may actually provide us with a compelling reason to reject it. As a result, considerations of self knowledge present no real obstacle to our finding an equilibrium point incorporating the ascriptional practices Burge appeals to.

3. Conclusion

It has been argued in this chapter that while the incorporation of the socially infected ascriptions that Burge appeals to into an equilibrium for our semantic terms is considerably more controversial than the incorporation of the ascriptions appealed to by Kripke and Putnam, such an incorporation remains preferable to trying to find an equilibrium that does not endorse the truth of such ascriptions. It was argued that such ascriptional practices are not only themselves fairly deeply entrenched but are also not clearly isolatable from the extremely well entrenched practices that Kripke appeals to. Attempts to dismiss or 'explain away' non-individualistic ascriptions by appealing to more 'pragmatic' factors were shown to rely upon a number of unrealistic assumptions about what we take ourselves to be doing when we make ascriptions or defer to

---

I do not have in mind everyday understanding of each other but, rather, what underlies everyday understanding or what everyday understanding consists in. The question, what it is for agents to have the concepts and contents which we attribute to them in our everyday understanding of them, cannot be answered by looking to the way in which we attribute those concepts and contents to them in everyday situations. (Bilgrami 1992a, 5-6.)

correction, and even those ascriptional practices that might have seemed to favor individualism (those associated with slips, malapropisms and misunderstandings) were shown to be best understood within an non-individualistic framework.

General beliefs associated with our speaking a shared language whose terms have determinate extensions favor incorporating the Burge phenomena, and it has been argued that the most frequently cited candidates for entrenched general beliefs incompatible with non-individualistic ascriptions (those relating to self-knowledge and behavioral explanation) turn out to be compatible with them after all. Indeed, accounting for self-knowledge ultimately turns out to cause more problems for the individualist. Reconciling our ascriptional practices with these general beliefs has involved picking up on some of the themes introduced in chapter two by stressing (1) that the practices are *non*-individualistic rather than *anti*-individualistic, (2) that the *de dicto*/*de re* distinction is a distinction between types of ascriptions, not between types of beliefs, (3) that the position is *methodologically,* rather than *ascriptionally*, individualistic, and (4) that the position relies heavily on our *self*-interpretations to fund the non-individualistic ascriptions. (Considerations similar to these four will turn up again when we discuss the possibility of incorporating the Wilson ascriptions in the next chapter.) There seems good reason to think that such strategies could also be used to reconcile the practices with any other seriously entrenched general belief. As a result, while the prospects for finding an ascriptionally individualistic equilibrium seem unpromising, there does not seem to be any reason to think that we cannot find an ascriptionally non-individualistic equilibrium for our semantic terms.

<u>Chapter Four: Linguistic Norms and Future Behavior</u>

1. Introduction

The purpose of the previous chapter was to show that the ascriptional practices that Burge appeals to can and should be incorporated into an equilibrium for our semantic terms. The purpose of this chapter is to argue that, once one accepts the admissibility of the ascriptional practices that Burge appeals to, there is little reason not to accept the 'retrospective' attributions of content that we commonly make. Once our general beliefs about meaning are brought into line with the ascriptional practices appealed to by Kripke, Putnam and Burge, there isn't much left to prevent us from incorporating the ascriptional practices brought to our attention by Wilson. Most of the general beliefs that conflict with such ascriptions will have been given up or modified already. Indeed, it will be argued in section five that there is a sense in which the practices appealed to by Putnam just *are* the ones appealed to by Wilson, with the true nature of those practices only being made clear with the Wilson cases.

Incorporating such future directed ascriptional practices into any equilibrium for our semantic terms involves accepting a type of "temporal externalism" (hereafter "T-externalism") according to which the future behavior of an individual or his society can affect the conditions for the correct application of his terms. This chapter will discuss, motivate, and elaborate various aspects of a T-externalistic position.

Before beginning, however, I should make two terminological points. First of all, when discussing T-externalism, talk about subsequent behavior "helping determine the term's final extension," or future use "making our concepts more determinate," is hard to avoid. Nevertheless, it invites a misunderstanding of the proposed view. Namely, it suggests that our concepts change over time (that is, they go from being indeterminate to being determinate), rather than having the same determinate content attributed at all times. According to the T-externalist, a term's meaning is determined by the totality of facts relating to its use, and such facts are not limited to what is epistemically accessible to us in

the present. Both past *and* future usage is relevant to the term's interpretation. Future behavior helps determine a term's extension, but it does so by constituting some of the facts relevant to its proper interpretation. Before the future decisions were made, the extension was not indeterminate. Rather, the extension was just 'unsettled,' since all of the relevant facts were not realized, or epistemically accessible. Future use can "make a term's extension more determinate" in an epistemic sense, just as adding a piece to a puzzle can make what the puzzle represents 'more determinate' without in any way changing what the puzzle represents.

Secondly, it should be noted that, in light of the views presented in this chapter, "extension" could be understood as ambiguous between (1) what the term refers to, and (2) how the term's use is 'extended' into the future. Since it is argued that (2) helps determine (1), context alone may not necessarily make clear which of the two uses is intended. Because of this, I will try to use the term exclusively in the first of the two senses given.

2 T-externalism: the ascriptional practice

The fact that future usage can affect the way we at least *attribute* thought and linguistic content has been discussed in chapter one. The gist of these cases is that our everyday ascriptions of content reflect a sensitivity not only to the agent's internal make up, and the structure of his physical and social environment (including the past history of use of the terms he is using), but also to 'accidental' developments in his terms' usage in the period subsequent to the moment of utterance. The cases in question will be recapitulated and expanded immediately below.

2.1 Wilson's Druids

For instance, imagine a speaker, Edwin, who is a member of an isolated community inhabiting an island on which the class of birds and the class of flying things are coextensive. He has a term "ave" whose putative extension includes these locally co-extensive classes, and he has beliefs such as "all and only aves can fly," and "all aves are

living things."[276] Both of these beliefs are true of the "aves" on the island, but it is clear that one must be given up if, say, a plane lands on the island and he encounters a flying thing that is not a bird. We can imagine that Edwin's total belief set will be split in just this way, with half of Edwin's ave-beliefs favoring one possible equilibrium, and half of his beliefs favoring the other. There may, however, be no way to tell in advance whether an equilibrium for Edwin's use of "ave" will include planes in its extension or not, because the mode in which Edwin first encounters planes may determine the comparative entrenchment of his beliefs and attributions.

While a term's putative extensions include everything one is disposed to apply it to, items to which the term has actually been applied tend to be more deeply entrenched.[277] For instance, if Edwin first sees planes flying high in the sky, his belief that all flying things are aves will lead him to call the planes aves. Planes may thus already be entrenched within the putative extension of "ave" when Edwin first sees one land and realizes that it is not a living thing. As a result, he may, upon seeing the first grounded plane, give up the belief that all aves are living things, and thus reach an equilibrium in which "ave" means *flying thing*. On the other hand, if he first sees planes on the ground, his belief that all aves are living things will entrench planes within the putative *anti*-extension of "ave."   This entrenchment may cause him, on discovering that planes can fly, to reject the belief that all flying things are aves, and thus reach an equilibrium in which "ave" means *bird*.

In each case, Edwin will not see himself as having changed what he meant by "ave" in any way. He will describe himself as having discovered that he had always been wrong in thinking that all aves were living things, or as having discovered that he had always been wrong in thinking that all flying things were aves. He will not see his use as just developing along one of two paths that were available for him. Instead, he will see his

---

[276]   We can assume that the rest of Edwin's community is in a similar position. Indeed, the example here is adapted from Wilson 1982, and Edwin is a member of Wilson's community of "Druids."

[277] One can thus reach a middle position between understanding the content of a speaker's utterances purely in terms of the *actual* causes of his behavior (as in Davidson's position), and one that seeks to understand content in terms of the *subjunctive* causes of that behavior (as in Fodor's position).

present use as being completely determined by his previous use. The order of appearance of novel phenomena may affect the comparative entrenchment of elements within the putative extension and general characterization, and thus determine which of a number of possible equilibria is reached.[278]

2.2 Grant's Zebra

Furthermore, how our present utterances are interpreted may be affected not only by our *own* future behavior, but also by the future behavior of *others* (behavior that may even occur after our death). To return to an example from chapter one, around 1820 Grant introduced the term "Grant's Zebra" for a type of zebra native to Kenya.[279] A few years later, Chapman introduced the term "Chapman's Zebra" for a morphologically distinct type of zebra found in present-day Zimbabwe. Later still it was discovered that the two types of zebra interbred near the Zambezi river and that, morphologically, one gradually faded into the other. Grant's and Chapman's zebras both turned out to be a races of the species *Equus burchilli* (one race of which, the Quagga, is arguably not a type of zebra at all).[280] This story seems representative of the way we typically describe such cases. However, if it is accurate, then (as with Edwin's use of "ave") it is merely a *historical accident* that the term "Grant's Zebra" had the extension it did. If the taxonomists had investigated the area around the Zambezi river *before* they hit deepest Zimbabwe, then they probably would have 'discovered' that Grant's Zebra could be found through most of East Africa, gradually

---

[278] In this respect, linguistic development is much like case law, which, notoriously, has this property. The principles according to which judges develop a set of precedents are often determined by the order in which new cases are brought to them. For a discussion of this aspect of case law, see Dworkin 1977.

[279] Of course, we do not know precisely what general beliefs Grant had at when he introduced the term, but we will assume here than his beliefs were equally compatible with either set of developments within the linguistic practice he initiated. In particular, even if he had some tendency to understand 'type of zebra' as 'species of zebra,' the fact that there are not, technically, any 'species level' classifications which include only zebras gives some incentive not to understand 'type of zebra' as 'species of zebra.'

[280] For a discussion of how the folk classification "zebra" does not seem to map on to any well defined biological kind see Gould 1983.

changing into a different subspecies as it drifted south. In such a case, "Grant's Zebra" would have picked out the entire species, not just the race found in Kenya.

2.3 Reaction to the practice

Such cases[281] suggest that, when we interpret the past use of other speakers (and even ourselves), we help ourselves to subsequent specifications that were not determined at the time of utterance. Many, however, are strongly inclined to think that, if they involve a commitment to T-externalism, such ascriptions *must* be mistaken. There is a temptation to think that what Edwin means by "ave" today *can't* depend upon whether the plane he sees tomorrow is in the sky or on the ground, and what Grant meant by "Grant's Zebra" in 1820 *can't* depend upon whether other naturalists explored Zimbabwe or the Zambezi in 1823. The conflict we see in the Wilson cases is one in which many, if not most, people are willing to say that the ascriptional practice is mistaken.[282] That is to say, they think that we should hold on to the belief that the content of our thoughts and utterances is independent of future linguistic development, even if it means giving up many of our ascriptions.

It should be stressed, however, that when just a single one of the scenarios is presented, the attribution of the later determinate content to the earlier stage seems perfectly natural. We are usually faced with just one such scenario when we actually go on to get a more determinate conception of what we are talking about. The intuition that makes this attribution so natural is, in some sense, tied to the presupposition that there will not be two alternate scenarios -- which is why the presentation of both scenarios undermines the intuition that is so strong in the single case. It is an important part of our linguistic practice

---

[281] Wilson discusses Grant's zebra, along with a number of other such cases, and argues that they occur "in virtually every case of enlargement of our world view through scientific progress." (Wilson, 1982, 572.)

[282] Including (possibly) Wilson himself. Wilson's case is tricky because he claims that in cases such as Grant's use of "Grant's Zebra" or Edwin's use of "ave" the reference of the terms in question changes but their meaning does not change. It is not, however, entirely clear that this is a consistent position, since it would allow that two successors of a given community could both mean the same thing by a particular word, even though they manifestly used it to pick out different classes of things.

that the people participating in it typically assume that how their terms are applied in novel cases will be determined by past use. What the Wilson cases suggest, however, is that this widespread assumption is often false. This is, of course, a uncomfortable realization if the practice described is one's own, and it might suggest that language is one of those practices the understanding of which undermines one's ability to take part in it. Furthermore, since cases where there are multiple equilibria available for our terms involve a breakdown of our presuppositions about language, any account of what our terms mean in such cases will seem unintuitive in at least some respects.[283] As a result, it is not obvious that there will be any completely satisfying way to describe cases where the putative extensions and general characterizations of our terms have multiple equilibria available to them.

As long as one understands speakers in terms of belief and truth (rather than remaining at the level of general characterizations and putative extensions), disallowing any contribution from future behavior leaves the extension of such terms indeterminate. On such accounts, what, say, Edwin initially meant by "ave" would be indeterminate between *bird* and *flying thing*, and the reference of his term would have become more determinate (and so, in a clear sense, changed) when he first encountered planes. The ultimate acceptability of a T-externalistic equilibrium will depend, to a large extent, on the comparative merits of such indeterminacy-based accounts. Such accounts are discussed in the seventh section of this chapter. While there may ultimately be no fact of the matter as to whether T-externalistic or indeterminacy-based accounts of content are to be preferred, it is clear that the latter have been more fully explored. The following few sections will focus on T-externalism and argue that this much neglected position, while perhaps not compulsory, is a surprisingly attractive option, perhaps even more attractive than its indeterminacy-involving competitors.

---

[283] We see a similar phenomenon with cases of splitting in discussions of personal identity: our conceptual apparatus works under the assumption that there will be no splits, and no application of this apparatus seems entirely satisfactory in the splitting cases. As Lewis puts it in another context, "Rival philosophical theories have their prices, which we seek to measure. But it's all too clear that for philosophers, at least, there ain't no such thing as a free lunch." (Lewis 1983a 353.)

3 T-externalism: arguments for the view

3.1 Specifying one's commitments, and the diachronic division of labor

As mentioned in chapter one, there is an argument form in which, given a set of criteria assumed adequate for determining what our words and thoughts are about, one constructs a thought experiment in which two individuals are identical with respect to the proposed set and yet can plausibly be said to refer to different things, thus showing that more must be added to the proposed set of criteria. Putnam is famous for arguing that our relations to the physical environment must be added, and Burge makes a similar point with respect to the social environment. The Wilson cases can be seen as suggesting that the future linguistic development should be included in the set of what contributes as well. Still, while it may seem as if the three factors represent a gradual expansion of the relevant factors, it should be noted that one can accept any of these additions independently of any of the others. In particular, one can accept that future events can determine what we mean, without accepting that our social context affects what we mean in a similar fashion. For instance, the ascriptions that tied what Edwin meant to his future usage could be accepted by someone who did not allow social usage to affect what an individual meant. The Grant ascriptions could not be accepted by such a person.

Indeed, the Edwin ascriptions might be the more compelling of the two for precisely this reason. When what we have said in the past underdetermines what we have committed ourselves to, it seems most plausible to say that at least *our own* subsequent utterances can be taken to specify the commitments involved in our earlier ones. One can be given credit or discredit based upon the correctness of one's assertions, and the assertions one goes on to make in the future can help determine the correctness conditions by which the entire set is evaluated. For instance, Edwin's initial use of "ave" had many aspects that favored interpreting him as meaning *flying thing* by the term, but just as many favoring interpreting him as meaning *bird* by it. As a result, Edwin could have found good reasons for claiming

that his term was properly used in either of the two ways. Given this background, we allow that Edwin always meant, say, *flying thing* by "ave" largely because he understood *himself* this way. If he had understood "ave" as picking out *birds*, we would have accepted his judgment on that as well. Unless our previous usage and dispositions clearly commit us to a certain equilibrium, we are entitled to specify just what we are committed to with our words.[284] The fact that Edwin could not, if pressed, find anything in his previous use or dispositions that justifies his choosing one equilibrium over another does not prevent him from having the right to do so. Philosophical self-exegesis is a clear case where this consciously happens,[285] but it often occurs without the speaker realizing that he is resolving a potential ambiguity. Even when there is more than one equilibrium available for our terms, certain contexts will make particular patterns of resolution seem so obvious that alternatives are not considered. Edwin's case has precisely this property. While two equilibria are available to him, the way in which he initially encounters his first planes will make the move towards one of these equilibria seem obvious.

However, once we allow that individuals have the right to specify their commitments this way, the fact that one can have one's reliance on others deeply entrenched in the general characterizations of one's terms can result in the equilibria for one's *own* terms being dependent upon the future behavior of *others*. For instance, if Edwin's later use can determine what he *always* meant by "ave," and there are people who rely upon Edwin for their use of "ave," then Edwin's future use can determine what *they* always referred to by "ave." If his neighbor picks up the term from Edwin, and died before Edwin first encounters planes, Edwin's usage after his friend's death could determine what that friend referred to by "ave" before dying.[286]

---

[284] For a fuller discussion of this, see the discussion of methodological individualism in chapter three.

[285] This phenomenon is discussed in Grice 1957, 222-3. [286] What we mean by our terms, just like whether we are *eudaimon* or not, is largely dependent upon what we do during our lives, but it is ultimately independent neither of what others do around us nor of events that take place after our death. For a discussion of how future events affect our happiness, see Aristotle 1985, 1100-1101.

Of course, we rarely have a particular expert in mind when we rely on community use. Rather, we understand ourselves as taking part in the various linguistic practices associated with our words. As discussed in chapter three, this belief is often deeply entrenched, and so our words are often correctly taken to have the semantic properties that the surrounding practices associate with them.[287] Such practices are not naturally understood as limited to thin temporal slices corresponding to the moment of utterance. Rather, the very idea of a linguistic practice involves some notion of temporal extension. Indeed, the admission of the Kripke phenomena and the contribution of past use already commits one to a notion of linguistic practices extending through time.

Just as an individual's usage does, a linguistic practice can have a number of equilibria available to it. The practice need not actually be viewed as *changing* unless it settles on an equilibrium that was not a member of the set originally accessible to it.[288] A practice can evolve as its characterization of a term's meaning is made more determinate, but as long as it is only made more determinate, it can still be understood as the same practice. If the practice remains the same, so does the meaning of the term dependent upon it. When we make an utterance, we often commit ourselves to future refinements in communal usage because these refinements determine just what linguistic practice our usage is (has always been) a part of. Our conception of our linguistic practice determines just what our reliance on that practice commits us to.[289]

It was argued in chapter three that an individual speaker can be held to expert use in part because of his assumption that he is talking about just what the experts are. In much the same way, the experts in a given society can be held to their successors' use in virtue of seeing their predecessors, their successors and themselves (and their successors' seeing them) as part of a continuous process of discovering the extension of the same terms.[290]

---

[287] For another discussion of this, see Evans 1982, 387.
[288] This topic is discussed further in chapter one.
[289] Once again, this sort of issue is discussed in greater detail in chapter three.
[290] "Darwin's claim that diversity of species is due to a process of natural selection … has undergone considerable amplification, qualification, and extension, but there are very few scientists today who

The synchronic division of labor is manifested in our confidence that, while we may not know exactly what falls under the extension of a given term, somebody else does. The diachronic division of labor is manifested in our confidence that, even if nobody knows exactly what falls under the extension of a given term, once somebody did, or someday somebody will. A scientist may introduce a term, knowing full well that its extension cannot be picked out determinately, leaving this task to future scientists (or perhaps his future self). The history of terms such as "AIDS," "hepatitis c," or "human growth factor" are typical examples of this process. By reading present specifications into the past, we license future interpreters to read future specifications into the present. Rejecting the correctness of such future ascriptions would commit one to the incorrectness of many of the ascriptions that one makes in the present, and since such future ascriptions are compatible with one's present utterance, there is no reason to take on such a commitment.

That there is a diachronic division of linguistic labor involving past usage is relatively uncontroversial, but it is a point that is all too easily forgotten. If what we meant had to be determined entirely by present usage, much of the name-using practice that Kripke drew our attention to would be ruled out. Present applications of a term can be understood as mistaken at least partially because what a speaker now means by a word he is using is typically understood as being a function not only of how he is presently using the term, but also of how he (and others) have used the term in the past. The claim that what we mean by our terms is not solely a function of *present* usage should, then, be fairly uncontroversial. Some might, however, be less willing to allow the contribution of future usage than they would that of past usage. After all, allowing past usage to determine what we presently mean would seem to involve forwards causation, which is fine, while allowing future usage to determine what we presently mean would seem to involve backwards causation, which is far from fine. This is, of course, a serious worry, and

---

would deny that *Darwin's* claim (and not just the more sophisticated selectionist positions which have since been developed) is true." (Rouse 1987, 152.)

sections 4 and 6 will contain a discussion of how T-externalism involves neither a commitment to backwards causation nor a commitment to any substantive theory of time.

Of course, the mere fact that there is a diachronic division of labor extending into the future does not entail T-externalism. If facts about present use already settled what we meant (i.e. if there were only one accessible equilibrium), then the diachronic division of labor could be understood as merely the case of the future generations bringing to light commitments that were already settled in the past. It is only once one realizes that past usage does not settle how future usage should develop that one can see how the claim that the future clarifications determine what we mean leads to T-externalism.

3.2 Determinacy

As discussed in chapter three, one's belief that one's terms have determinate extensions favors the contribution of one's present social environment. Since admitting future usage can make what would otherwise be indeterminate contents determinate, this same belief supports the contribution of future usage as well. Neither Edwin nor Grant understand themselves as meaning something indeterminate by their terms. However, unless future use is allowed to help determine the extension of our terms, their extension will often be indeterminate. The fact that an individual's, or even an entire society's, usage can, up to any given time, fail to pick out a determinate extension for an individual's term gives one reason for trying to expand the pool of evidence relevant to what the individual picks out with his terms. Given that we actually make ascriptions that help themselves to future specifications, future usage is an obvious candidate to be included within the expanded pool.

3.3 Shared languages

Just as it is an advantage of non-individualistic accounts of content that they accommodate the idea that members of a linguistic community mean the same things by their words, it is an advantage of T-externalistic accounts of content that they can

understand present and past stages of a linguistic community as talking about the same things with the same words. The continuity that we are preserving is not only between past and future stages of our society, but between our past and future selves.

Indeed, the T-externalistic position, by allowing for more continuity of content, allows for more continuity between our past, present and future psychological states. If one is inclined to think of personal identity in terms of psychological continuity, then the T-externalistic position is much friendlier to the thought that we preserve our identity through time.[291] In much the same way, one's having a substantial conception of preserving our identity though time will also make one more sympathetic to T-externalism. If one understood persons as successions of temporal slices that were merely more-or-less similar to each other, then it might seem that the commitments one such slice could undertake at a moment could not be affect, or be affected by, the actions of the other slices. On the other hand, if one understands people as preserving their identity through time, what one commits oneself to at a moment can be affected by the actions that one makes at other moments. If the locus of commitments is itself temporally extended, then the claim that its assertional commitments need not be settled at the moment of assertion will seem more plausible.

4 T-externalism: an elaboration of the view

4.1 Meanings and histories

Since T-externalism gives future behavior a significant role in determining what we presently mean, it should not be surprising that some of the formal apparatus designed to account for explicitly future directed statements can also help explicate a T-externalistic

---

[291] Assuming, of course, that one understands psychological continuity in terms of the preservation of 'wide' contentful states. 'Narrow' psychological continuity would be preserved equally well on either view. Continuity of narrow psychological states is, however, a less appealing criterion of personal identity than the preservation of more fully contentful states.

semantics. In particular, the Edwin and Grant cases will here be characterized in terms of the following:[292]

1. A *moment* ("*m*") is a spatially complete but instantaneous event: all of nature idealized to a zero temporal thickness.

2. The *causal ordering relation*, $m_1 \leq m_2$. (This relation is taken to be reflexive, transitive, and anti-symmetrical.)

3. A *history* ("*h*") is a maximal chain of moments.

4. *Our World*. The set of all moments that are connected to this very moment by means of any zigzag combinations of the causal ordering relation or its converse.

The structure of Our World has the following features:

5. Historical connection: every two moments of Our World have some common historical ancestor.

6. No Backwards Branching: all branching is forward, never backward. Incompatible moments in Our World never have a common upper bound.

A bare moment is not enough for the evaluation of the truth of all of the utterances made at that moment. What one needs is a moment/history pair. The truth of many of our utterances, especially those about the future, depends upon how history continues after the moment of utterance, and so such histories must be included in the context of evaluation. The truth of a sentence is always relative to a history.[293]

Utterances that have their truth-value established at a moment, independently of what happens afterwards, can be described in terms of "settled truth." Roughly, an assertion is *settled true* at a moment if it is true in *every* history passing through that moment.[294] In much the same way, a term's extension can be said to be 'settled' at a moment if the term has the same extension in every history passing through that moment.[295] This is only a rough and ready characterization of what it is for an term to have a settled extension, since

---

it assumes that, by the end of each history, each term's extension will have been settled. There will, however, be histories where a term's usage dies out before it is fully specified, and to account for such cases, one will need to modify the account of what it is for a term to have a settled extension as follows. A term will have a settled extension at a moment if either (1) there is at least one history containing that moment in which an equilibrium is reached and in every history containing that moment where an equilibrium is reached, the same equilibrium is reached (this should account for moments where a term has a settled extension even though there are some histories containing it in which no equilibrium is reached), or (2) the extension of the term in question has already been settled at some previous moment in all the histories containing that moment. The meanings of many of our terms may be 'settled' in just this fashion. In spite of this fact, one should not equate 'real extension' with this notion of 'settled extension' (which would suggest that if something is without a settled extension at m_i, then it doesn't 'really' have an extension at all). 'Real' extensions are always relative to a history.[296]

While the contribution of the future in determining the truth-value of statements either in the future tense or containing temporally loaded expressions such as "the losing bet" or "the final battle" has long been recognized, the T-externalist suggests that the class of these statements will be much larger than initially thought. In particular, the contribution of the future is often needed not only to determine the truth of the proposition expressed, but also to determine just what proposition *is* expressed. It should be noted, however, that even if the interpretation of a speaker's words may not be settled at a particular moment, the truth value of his utterance may be. Even if what Grant meant by "Grant's Zebra" depended upon how future usage developed, his utterance of "My zebra has stripes" will be settled true when uttered, since it is true on either candidate interpretation. This will not, however, always be the case. For instance, the zebras in Zimbabwe do not have striped feet, but

---

[296] Belnap and Green (Belnap and Green 1994, 374) argue against equating 'real' with 'settled' truth, and if one sees a close relation between truth and meaning, one will expect a structural analogy here. If a term doesn't have an 'settled' extension at m_1, it should not be assumed that it is then 'settled' that it has an indeterminate extension. [297] Lewis 1974, 110.

those in Kenya do. As a result, the truth value of Grant's utterance in 1820 of "My zebra has striped feet" will vary from history to history, and thus be unsettled at the time of utterance. However, it would be misleading to see future linguistic behavior as something that *makes* Grant's claim true or false. Future usage determines the utterance's content, not its truth value. It makes Grant's claim the claim it is, it does not make the claim true. What makes the claim true is still the fact that Grant's Zebras have striped feet.

According to the T-externalist, then, interpretations are relative to histories, and the context that determines the proposition expressed by an utterance will include both past and future usage. Nevertheless, just what consequences T-externalism has for what we mean by our terms will depend partially upon how we understand the future. The effect on our picture of content of allowing 'future facts' to help determine what we are talking and thinking about depends upon just what our conception of these 'future facts' is. We should thus discuss T-externalism in relation to three conceptions of the future: determinism, indeterminism with the so-called 'thin red line,' and the 'open future.' The last of these will be discussed in greatest detail, since it is the conception of the future under which T-externalism is the most unintuitive.

4.2 Determinism

Determinists believe that there is only way in which the future can possibly develop. A determinist will thus understand the passage of time from the past, through the moment of utterance, and into the future as consisting of a single history (h_1) which can be represented as follows:
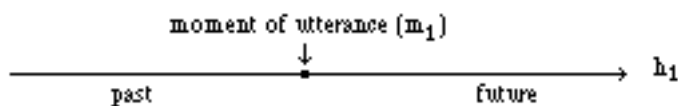

If there is only one history, then every truth will be true in every history, and thus all truths will be settled truths. As a result, if determinism is true, then it has *always* been settled that, say, the zebra enthusiasts would hit deepest Zimbabwe before the Zambezi river, and

thus *always* settled that the term "Grant's Zebra" would be applied to just the zebras in Kenya. A determinist thus need not see the Wilson cases as presenting a type of semantic indeterminacy at all. Such indeterminacy was supposed to arise because there were two alternate ways in which the usage could develop, but if determinism is true, there is no real alternative to our actual development. The future contains a set of settled facts that, while epistemically inaccessible to us at present, remain settled facts nevertheless. Allowing future events to contribute to the present meanings of our terms would thus not prevent us from taking there always to be settled facts about what we mean. It would thus seem less unintuitive to say that future events contribute to the present meanings of our terms. Some of the relevant facts may be epistemically inaccessible at the moment of utterance, but they remain settled facts nevertheless. A deterministic T-externalist can follow Lewis in insisting that we should not be interested in how *we* determine the facts, but rather in how *the facts* determine the facts.[297]

4.3 The thin red line

Much the same can be said if one accepts indeterminism with the thin red line (hereafter TRL). Indeterminism involves there being just one past, but a number of possible futures (represented here by a tree structure with a common trunk leading up to the present and various branches extending from it into the future). TRL assumes that, while there are a number of ways in which the future *could* possibly develop, there is still a fact of the matter as to what *will* happen. The possible future that is also the actual future can be understood as a "thin red line" (represented here as a thick black line) passing through the tree representing the possible histories.

Possible futures: h_1, h_2, h_3, h_4
Actual future: h_3
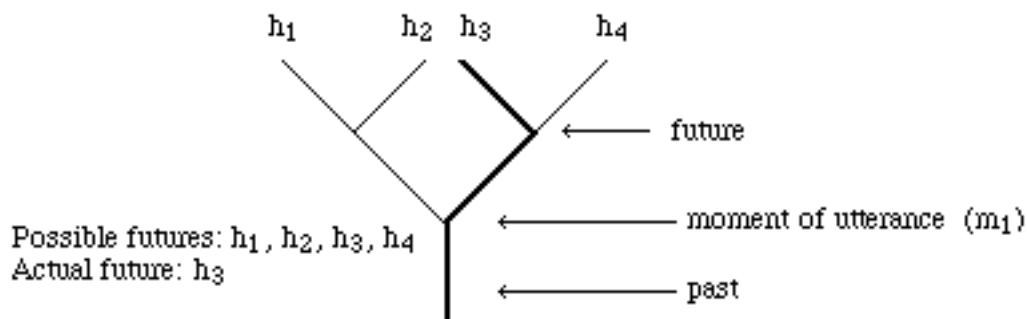
The T-externalist can describe the Grant's Zebra case with such a picture of the future as follows. Let us assume that Chapman discovers the zebras in Zimbabwe before explorers find them on the Zambezi on h_1 and h_3, while the order of discovery is reversed on h_2 and h_4. On such a picture, we can say that, at m_1, while it remains *possible* for the explorers to investigate the zebras by the Zambezi first, they will, *in fact*, see ones in Zimbabwe first. With TRL, the admission of indeterminism does not prevent there from being a fact at m_1 about what will happen after it. The notion of epistemic inaccessibility may be stronger here, since with determinism it is at least in principle possible to discover how future usage will develop by looking at how things stand at m_1. If determinism were true, knowing the states of all the sub-atomic particles at m_1 and the laws of physics would allow one to know, in principle, how the usage will develop. A Laplacean demon could always calculate future use based upon its knowledge of present facts. No such calculation could be made with TRL. Still, in many important respects, TRL is like determinism. While it is not a fact *about* m_1 that zoologists will not go on to apply "Grant's Zebra" to all of *Equus burchilli*, it is still a fact *at* m_1 that they will not do so. By contrast, on the deterministic picture, it is both a fact *about* m_1 that they will not go on to apply "Grant's Zebra" to Zimbabwean zebras, and a fact *at* m_1 that they will not do so.

With both determinism and TRL, the future events just reveal what always was to be. All the facts constitutive of the meaning were already in play, and thus so was the meaning itself. The passing of time simply makes open to us facts that were previously epistemically inaccessible. This makes the ascriptions associated with Wilson's thought

experiment more palatable. If it was already true at m_1 that Chapman would (in fact) discover the Zimbabwean zebras, before any zoologist hit the Zambezi, then it seems less strange to say that Grant meant just Grant's Zebra (and not *Equus burchilli*) by "Grant's Zebra" even at m_1.

4.4 The open future

On the other hand, if the future is "open," then, at the time of Grant's initial dubbing, there will be not facts about how usage will develop, and thus about which equilibrium will be reached. A T-externalist who took the future to be open would thus have to describe the Grant's Zebra case as follows:
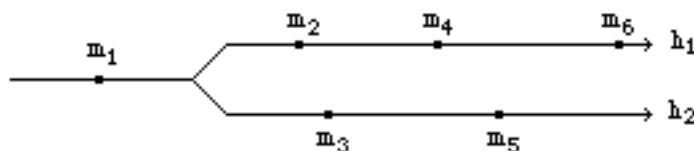
Treating m_1 as Grant's initial use of the term, there are two possible histories, h_1 and h_2, in which the use of the term develops in different ways. In *our* history, h_1, the term "Chapman's Zebra" was introduced at m_2, the interbreeding discovered at m_4, and we are discussing Grant's utterance at m_6. In the alternative history, h_2, the gradual shift at the Zambezi was discovered at m_3 while the Zimbabwean zebras were found at m_5. On h_2, "Grant's Zebra" denotes all of *Equus burchilli*, while on h_1 it denotes just Grant's Zebra.

T-externalism with the open future thus leaves it at m_1 not only *epistemically* but also *metaphysically* unsettled what Grant means. It is neither a fact *about m_1* nor a fact *at m_1* that the "Grant's Zebra" using practice will develop in any particular way. However, while some might find this fact a little 'unsettling,' it seems irrelevant to our *present* interpretation of Grant's 1820 utterances containing "Grant's Zebra." There are *now* settled facts about Grant's society's later history just as there are settled facts about its prior history. While it may be the case that, at the time of Grant's utterance, it wasn't *settled* that this would be the history that became the actual one, relative to *this* history, h_1, Grant's term *always* meant

Grant's Zebra. (I.e., it is now settled that he meant Grant's Zebra at m_1, though it was not settled at m_1 that he meant Grant's Zebra.) At m_1 the full context of evaluation was not in place. Furthermore if there really is an open future, then evaluations of our utterances will have to be relativized to possible histories *anyway*. We would look back on Grant's future tensed utterance of "I will never be discussed in anyone's dissertation" and say that it was false, even though, at the time of utterance, it was not settled false. In an analogous fashion, we should look back at his utterances involving "Grant's Zebra" and treat them as referring to Grant's Zebra, even though, at the time of utterance, that extension was not yet settled.

5 T-externalism: three further consequences

There are a number of interesting consequences and incidental payoffs that result from the adoption of a T-externalistic position. Issues that seem puzzling within a temporally bound framework can lose much of their paradoxical character once future use is brought into play. Three such issues will be discussed briefly here. Once one allows future use to contribute towards what speakers meant in the past, the so-called "causal theory of reference" for natural kind terms can be stripped of some of its more dubious metaphysical and psychological presuppositions.[298] We can say that we have always meant what we now do by terms such as "water" or "gold" without having to say that there is some special connection between linguistic usage and natural kinds. The initial baptism need involve neither unrealistic 'referential intentions' on the baptizer's part nor the possession of metaphysically privileged 'elite properties' on the part of the object baptized.[299] Our linguistic precursors did not pick out gold by their term corresponding to "gold" because mineral kinds like gold are somehow 'more real' than functional kinds such as "yellow metal." Our precursors almost certainly had no explicit intention to pick out a natural (rather than, say, a functional) kind when they introduced the term. However, their vague

---

[298] For a discussion of some of the problems with such presuppositions, see appendix A.
[299] The first of these is appealed to in Putnam 1981d and Kripke 1972, the latter in Lewis 1983a.

presuppositions about the term probably made equilibria corresponding to both natural and functional kinds accessible. The compatibility of their usage with a number of equilibria allowed us to apply our own specifications retroactively. T-externalism allows one to incorporate an account of natural kind terms within a diachronic social picture, thus allowing one to provide a unified account of both the Putnam and the Burge phenomena.

T-externalism also lends itself to a fairly quick and cheap response to the skeptical paradox Kripke attributes to Wittgenstein. Kripke's skeptic assumes that if I meant something in the past, then it must be in virtue of some *past* fact about me.[300] The skeptic must initially admit that we *now* mean *plus* and not *quus* ("otherwise, we will be unable to *formulate* our problem"),[301] and the T-externalist can say that *this* is precisely what makes it the case that we meant *plus* in the *past*. Of course, this present stipulation would do no good if it just amounted to my present utterance of the noise "in the past I meant plus," since such an utterance would, itself, be subject to interpretation. However, given that present use is not in question, one can help oneself to fully contentful use in the clarification of what one meant in the past. When Edwin initially grasped the rule for "ave," his intention and the rule he grasped did "determine a unique answer for indefinitely many new cases in the future."[302] However, it was only settled just what this rule was once he became acquainted with airplanes. Past intentions do 'determine' correct answers for new cases in the future, but the determination here is more 'logical' than causal. The content of past intention is determined by the answers we give in the novel cases.[303] There can be, as it were, an 'internal connection' between how the speaker applies the term in novel cases and the rule he is taken to have grasped in the past.

---

[300] Kripke 1982, 13.
[301] Kripke 1982, 14.
[302] Kripke 1982, 8.
[303] Brandom (1989) argues that Hegel defends a similar view about the relation between our intentions and actions; that is, the content of the intentions we originally form is determined by the actions that those intentions result in. Just as becoming clearer about what one's practical commitments are as one acts need not be understood as a change of mind, clarifying one's theoretical commitments need not constitute a change of meaning as long as one stays within the range of accessible equilibria.

Still, while Kripke's arguments cannot show the T-externalist that I didn't mean *plus* by "plus" in the past, they could, if good, show that it wasn't *settled* that I did. One does not want to say that, if I had just recently given "5" as the solution to "68 + 57," then *quus* would have been what I *always* meant by "plus." The quick and cheap response alone would not prevent each new application of a term from being an "unjustified leap in the dark."[304] The fact that such leaps in the dark 'retroactively' make it the case that they are in accord with the rule does not change the fact that, if Kripke were right, previous use would not 'guide' or 'constrain' us with respect to novel cases. While this can *sometimes* be true (as in Edwin's use of "ave"), it is not *always* the case. One wants, then, to be able to say that while Edwin's use of "ave" may leave it open whether the term applies to planes or not, there does not exist a similar choice as to whether the term applies to, say, robins (or any other species of flying bird not previously encountered on the island). Appealing to the need to reach an accessible equilibrium (of the sort discussed in chapter one) between his term's putative extensions and general characterizations will be enough to do just that.[305]

Finally, T-externalism points towards a substantive version of the Pragmatist view that truth is "the opinion which is fated to be ultimately agreed to by all who investigate."[306] Peirce's maxim has often been criticized for lapsing into either falsity or emptiness once the notion of the opinion "fated to be agreed upon" is probed further. Nevertheless, a substantive, and still plausible reading of Peirce's maxim can be given in light of the considerations above. Namely, it can be understood as a conclusion about meaning, and not particularly about truth *per se*.[307] The meanings of our terms become settled through the process of inquiry, then one could plausibly say that the 'end of inquiry' is the ideal limit where an equilibrium is reached for all the relevant terms. Until this point is reached, the truth values of many sentences will remain unsettled. Furthermore, since (especially

---

[304] Kripke 1982, 10, 15.
[305] For a discussion of how this could be done, see my "We Live Forwards but Understand Backwards" (MS 1995).
[306] Peirce 1878, 273.
[307] William James' forward looking account of intentionality is especially hospitable to T-externalism. (See my "James' Pragmatic Theory of Intentionality" MS 1995.)

with the open future) there is no normative constraint that completely determines what the end of inquiry will be, the claim made is a substantive one. On different histories, inquiry will reach different ends and claims such as "Grant's Zebra has striped feet" will turn out to have different truth values. Assertional commitments are determined by the process of inquiry, and so truth-conditions depend upon the end of inquiry.

There are thus a number of advantages to adopting a T-externalistic framework. There may, however, be a number of disadvantages as well, and such potential problems are the topic of the next section.

6 T-externalism: some objections and replies

This section will consider, and attempt to answer, a number of objections to the suggestion that future usage can contribute to the present content of our thoughts and utterances.

Objection: T-externalism seems to endorse a type of backwards causation by allowing future events to determine the content of what was thought and said in the past.

Reply: While T-externalism allows future events to determine what we meant in the past, the type of determination involved is not causal. As a result, no backwards causation is involved. According to the T-externalist, the fact that the practice within which a given word is used extends through time allows future usage to help constitute the very practice that the speaker was taking part in at the moment of utterance. The speaker's utterance is understood in terms of a characterization of the practice that may not have been available at the moment of utterance, but the characterization remains correct nevertheless. T-externalistic characterizations no more involve backwards causation than do claims like "John bought the winning lottery ticket on Tuesday." John's ticket may only have become "the winning ticket" when his number was drawn five days later, but one does not require

any notion of backwards causation to claim that he bought the winning ticket on Tuesday.[308]

Objection: If T-externalism does not understand the relation between present contents and future use as a causal one, this itself presents a problem for this view. Externalist theories such as Davidson's perceptual externalism exploit precisely this causal connection between thought contents and the 'external' factors that are meant to affect it. If future usage has no causal affect upon present utterances, then it should not effect the content of these utterances.

Reply: While T-externalism does not involve any backwards causation, it doesn't mean that we cannot understand there to be any causal relations between present utterances and future usage. There is still a causal relation between the two. Not only can present utterances causally affect future behavior, but the past use that affects our present utterances can be described as being part of a practice that also includes future usage. We can thus say that while future usage does not causally affect our present utterances, it can help determine the relevant characterization of the past usage which *does* causally affect our utterances.

Objection: Unlike other types of externalism, T-externalism seems vulnerable to a paradox similar to one associated with the idea that courts can fill in 'gaps' in the law *ex post facto.* For instance, imagine that, in 1980, a couple decides to get married while on a cruise ship. The ship's captain starts the wedding ceremony when they are still at sea, but by the time the ceremony is over, the ship is at port. The upon disembarking, the putative husband is killed by a passing car and, in a battle over his estate, his relatives contest the

---

[308] Worries about backwards causation and T-externalism may bear some resemblance to worries arising from Aristotle's discussion of how we could be subject to "posthumous harm" (Aristotle 1985, 1101). In much the same way, the claim that we are subject to posthumous harm should not be understood in causal terms. Rather, it seems better understood in terms of future events making available descriptions of our life that make it seem less choiceworthy.

legality of his marriage. The law clearly states that ship's captains have the right to marry a couple at sea, and that they have no such right while at port, but it says nothing about whether they have such a right if the ceremony is begun at sea and ended at port. The case eventually reaches the state court of appeals in 1992, and the court decides that the marriage was, in fact, legitimate. If the courts count as filling in such gaps in the law *ex post facto*, then we can say that court *made it the case* that the couple was married in 1980. This seems to lead, however, to a paradox. If they were married in 1980, then, by the time the court made its decision in 1992, the couple had already been married for twelve years, and if they had already been married for twelve years, how could the court then "make it the case" that they were married. How can one *make* the case something that already *has been* the case for over ten years? One shouldn't be able to *make* something so if it already *was* so.

A similar paradox seems to arise for the T-externalist. Suppose that Edwin sees his first plane in 1990 at the age of thirty three, and that his use of "ave" between 1990 and 1991 makes it the case that he means *bird* by "ave." The T-externalist also seems committed to saying that his usage also makes it the case that he *always* meant *bird* by "ave." However, if Edwin *always* meant *bird* by "ave," then he meant *bird* by "ave" in 1970, and so by the time he had seen his first plane, he had meant *bird* by "ave" for over twenty years. How, then, can what Edwin did in *1990* make it the case that he meant *bird* by "ave" if he had already meant *bird* by "ave" for over twenty years?

Reply: Such paradoxes arise to a large extent because of the comparatively course-grained character of the phrase "make it the case that." It does seem paradoxical to claim that one made something the case at time *t* if it already *was* the case *before* time *t*. Such paradoxes disappear, however, if we describe Edwin's usage in 1990 as *settling it* that he meant *bird* by "ave," rather than *making it the case* that he did so. Talk of settling makes implicit reference to other possible histories, and so does not involve the same sort of commitment to what could not have been the case in the history in question. If Edwin

made it the case in 1990 that he meant *bird* by "ave," then it could *not* have been the case that he meant *bird* by "ave" in 1989. However, if he only *settled* it that he meant bird by "ave" in 1990, it could still be the case that he meant *bird* by "ave" in 1989. It just couldn't have been *settled* that he did. That is to say, there must have been possible histories passing though the moment in 1989 in which he did not mean *bird* by "ave."

Describing what Edwin did in 1990 in terms of settling it that he meant *bird* thus does not produce the paradox associated with claiming that he made it the case that he meant *bird*. Nevertheless, isn't the T-externalist still committed to saying that what Edwin did in 1990 made it the case that he meant *bird* by "ave" in 1960? It is not simply enough to find an alternative way of describing the cases which does not produce the paradox, the T-externalist must also show that he is not, in fact, committed to the style of description that does produce the paradox. Does the T-externalist have the resources to *deny* that what Edwin does in 1990 makes it the case that he always meant *bird* by "ave"?

To put himself in a position to make such a denial, the T-externalist needs to give an account of what *does* make it the case that Edwin means what he means, and then show how such an account prevents one from claiming that what Edwin did in 1990 made it the case that he always meant *bird* by "ave." The T-externalist is, in fact, positioned to do just this. According to the T-externalist, what Edwin means at a given time by a term is determined by his entire temporally extended practice of using that term. As a result, what Edwin means by "ave" in both 1970 or 1990 is determined by the entirety of his usage from 1960 through 1992. What makes it the case that he means *bird* by "ave" from 1960 through 1992 is his entire history of using the term from 1960 through 1992. His 1990 usage is certainly part of that history, but it is not by itself something that makes it the case that he meant *bird* by "ave" in 1970. It *is* the case that he meant *bird* by "ave" in 1970, but what makes this the case is the fact that he meant *bird* by "ave" from 1960 through 1992. Stretches of usage make it the case that one means something by a term for stretches of

time, but moments of usage do not serve in explanations of what one means for moments of time.

Objection: What one means cannot extend beyond what one could communicate, so information that is unavailable to one's interpreters cannot contribute to what one means. Future use is unavailable to one's interpreters, so it can't contribute to meaning.[309]

Reply: First of all, once could argue that there is a sense of "communicate" in which one can say that what was *communicated* at a time was also determined by future usage. After all, if Edwin tells a friend that he killed and ave on his twenty first birthday, then that friend can come to believe that Edwin killed an ave on his twenty first birthday. According to the T-externalist, the content of both Edwin's utterance and his friends belief will be determined by future usage, and so it is not much of a stretch to claim that what Edwin communicated by his utterance was so determined as well. The objection thus rests on a particularly restrictive notion of meaning an communication by which both are limited to what could be determined by a contemporaneous interpreter.

However, there is little reason to restrict the evidence for an interpretation to what is available to a contemporaneous interpreter, since doing so would exclude important features of the society's *past* use as well. Such a restriction would thus rule out much of the name-using practice that Kripke describes along with the T-externalistic ascriptions. Since theories limiting the scope of what we could mean to the epistemic horizons of a contemporaneous interpreter are not nearly as well entrenched as the backwards-looking ascriptions they conflict with, it seems fairly clear that one should give up the restrictive theory. The inability of an interpreter to determine what a person means given all the facts

---

[309] Something like this reason might be what leads Wilson not to draw T-externalistic conclusions from the druid and zebra scenarios. Wilson takes it to be a "rough adequacy condition" on any theory of meaning that "The evidence for assignment of an extension to a predicate should be limited to such linguistic behavior as can be reasonably extrapolated from the community's contemporaneous practice and should not reflect accidental features of the society's later history." (Wilson 1982, 553.) It is certainly true that, if we accept the first conjunct of Wilson's condition, we must accept the second. However, for reasons given in the reply, the first conjunct cannot be accepted.

about him at a given time should be taken as showing that all the "semantically relevant facts" need not be available at that time.[310]

One might, of course, try to accommodate the Kripke ascriptions by allowing past and present, but not future, usage to contribute to what we mean.[311] Nevertheless, one should have some principled reason for doing so, and once it is clear that causal influence isn't the issue here, the clearest motivation for this 'asymmetry thesis' disappears. Principled reasons of a quasi-internalist sort could be given for ruling out both, but it is much less clear how one is to justify ruling out one without the other. (Though reasons of a vaguely verificationist sort might be marshaled in favor of looking at the contribution of the future in preference to that of the past.) For instance, many of the intuitions that support the contributions of past usage (e.g. those of determinacy and continuity of meaning) support the contributions of future usage as well.

One might think that one could rule out the contribution of future usage because, if the future is open, facts about future use are not even *in principle* available to the interpreter. However, when we *now* interpret, say, Grant's original utterance, it is patently false that we don't have access to his terms' subsequent usage. What information is available to interpreters depends upon which temporal position one takes them to be interpreting from. The claim that what we mean must in principle be available to an interpreter has some plausibility, since what we mean by our terms is a quite plausible candidate for something which should not extend beyond what can be known. However, such arguments tying a range of facts to what can be discovered lose most of their plausibility if accompanied by such *ad hoc* restrictions upon the investigator's evidential base. As long as one allows an evidential base which includes future behavior, T-externalism is perfectly compatible with the requirement that what a speaker means by his terms be available to an interpreter. Finally, it should be noted that, future usage *is* in principle available to one's interpreter's,

---

[310] For a discussion of this, see Grandy 1973, 450.
[311] Indeed, one might think that this was Wilson's real intent when he proposed his adequacy condition. However, Wilson's positive account of linguistic meaning does, in fact, rule out any contribution from past use.

they need only wait. On the other hand, barring the possibility of time travel, there may be no way for our interpreters to become acquainted with crucial aspects of past usage. As a result, future usage may be better off than past usage when we consider in principle availability to interpreters.

Objection: We don't know how future use will develop, so if such use contributes to what we presently mean, then we don't know the content of our own thoughts and utterances.

Reply: Given that we are not authoritative about the world's physical structure or the usage of our community, temporally-infected ascriptions no more threaten self-knowledge than the ascriptions Burge and Putnam appeal to. Self-knowledge creates no problems that are unique to T-externalism, and the strategies discussed in chapter three for reconciling self-knowledge with the Putnam/Burge phenomena will do so for T-externalistic phenomena as well.

Of course, the inaccessibility involved is of a different sort with future use, since one might think that there is, at least in principle, a possibility of the agent finding out what he means in the Putnam or the Burge cases. For the Burge cases the speaker could consult a dictionary or ask an expert, while for Putnam cases such as our use of "water" in 1700, one might think that the speaker could, in principle, perform the experiments and come up with the scientific theory needed to discover the true extension of "water." However, this difference is not one that makes a difference when what we are concerned with is self-knowledge. The relevant conception of self-knowledge is tied to notions like introspection, not to our ability to, say, go look up a definition in a dictionary. As a result, preserving self-knowledge gives us no new reason for rejecting T-externalism. Indeed, once again, given the practical inaccessibility of the past and the eventual arrival of the future, issues relating to self-knowledge should seem more pressing for those who allow the contribution of past usage than for those who allow the contribution of the future.

Objection: By allowing future use to determine the contents of our thoughts, T-externalism prevents us from making sense of the fact that our attitudes can be used to explain our behavior. All the facts having to do with whatever causally affects our behavior will have to be settled before that behavior occurs, so if T-externalism entails that the contents of our thoughts are not settled at the time of our action, then it seems to entail that our thoughts to do not causally affect our behavior.

Reply: If T-externalism does spoil the role of attitudes in behavioral explanations, then they were spoilt already by the practices Burge and Putnam describe.[312] As a result, if one accepts Burge's thesis about the social character of meaning, then problems with behavioral explanation will give one no reason to treat the practice Wilson describes as mistaken.

It should also be noted that T-externalism ascriptions can do a better job of 'making sense' of the speaker's behavior in the novel cases. We can explain why Edwin calls, say, the planes he initially sees in the sky "aves" by the fact that he means *flying thing* by "ave." By preserving the 'internal connection' between content and future applications, one can understand the speaker's behavior in novel situations as rule-governed in a way that one would not be able to otherwise.

Objection: By allowing future use to determine what we mean, T-externalism allows future developments that we would intuitively take to be *departures* from our use to determine what we 'really' meant all along. There seem to be clear cases where future development in a term's use should *not* be retroactively applied to the past. For instance, the term "girl" was once applied to both girls and boys, but one should not have to say that the fact that we now apply the term to just girls entails that previous applications of the term to boys were mistaken.

---

[312] Once again, this is a subject discussed in greater detail in chapter three.

Reply: Before considering this objection, it should be remembered that an equilibrium is "accessible" if it does not require giving up elements that are too deeply entrenched.[313] Not all possible equilibria will be accessible.   A term's usage at a time may leave open a number of equilibria accessible to it, and these accessible equilibria set limits upon how much linguistic usage can change without there being a change of meaning. In order for it to be read back retroactively, future usage must develop within the range of equilibria that present usage has accessible to it.   This severely constrains the types of linguistic development that can be read back into present use, and most perceived departures from present use will not satisfy such a constraint. If the changes in usage move to something that was not originally an accessible equilibrium, then there will be a change in meaning (as in the case of "boy"). On the other hand, if the change in usage can be understood as a move towards one of the accessible equilibria, there does not need to be any change in meaning (as in Edwin's use of "ave").

It is doubtful that any fixed procedure is available to determine when future use changes rather than clarifies what was meant before. Nevertheless, there should be a few rules of thumb about when a development is moving within the range of the term's accessible equilibria and when it is not. Finding these rules of thumb should illustrate that our pre-reflective notion of 'change of meaning' does correspond to a recognizable class of cases within the framework sketched in chapter one.

Changes in a term's conditions of application do not necessarily constitute a change of meaning. However, if enough of the original samples drop out of the putative extension (as in all the boys no longer being called "girls"), or if enough items from the term's putative anti-extension (those items originally seen as definitely not falling under its extension) are added (beads not part of a rosary coming to be called "beads") then, unless such radical shifts are justified by appealing to deeply entrenched aspects of the general characterizations, there may be a change of meaning.

---

[313] This is discussed in further detail in chapter one.

Radical changes in the general beliefs can occasionally cause changes in meaning as well. If we were suddenly to treat "water" as a functional kind term, we might be viewed as changing its meaning, even if the putative extension of the term remained the same (i.e.: if all the liquids we knew that had the required functional properties were made up of H_2O). However, it would constitute a change in meaning at least in part because the framework that was given up was a 'viable' one for "water." (That is to say, there is at least one accessible equilibrium that incorporates the natural kind framework.) If investigation had proved that water was not predominantly a single kind, and we could no more find a natural kind corresponding to "water" than we could for '"tree" or "lily" or "air,"[314] then we could reclassify it as a functional kind without changing its meaning.

Such an account would allow for considerable continuity of meaning through many radical shifts in our conception of the type of things we were talking about, but not all.[315] People who classified the animals in their environment in terms of a totemistic framework (members of certain tribes are descended from certain animals etc.) need not mean anything different than we do by their animal terms. This is because their most general, or 'framework,' assumptions no more determine what they mean by "goat" than our assumption that lilies formed a natural kind determined what we meant by "lily." There is no accessible equilibrium incorporating the totemistic framework, so giving up that framework need not constitute a change of meaning for one's terms. Things could be quite different with a people who classified things exclusively by their functional or aesthetic qualities (say, classifying birds by the type of meat they provided or the type of call they had). If they did not see the functional or aesthetic qualities as corresponding 1:1 with something like biological kind (or at least were willing to favor the functional classification if a conflict became apparent), then, even if they could be brought over to see the natural world in terms of 'natural kinds,' this conversion could be viewed as changing the

---

[314] For a discussion of how frequently our 'folk-biological' kinds fail to map on to any biological kind, see Dupré 1981.

[315] For another discussion of this, see the section on holism in chapter one.

meanings of all their terms, since the framework given up was a viable one. In any case, meaning change is not a problem unique to T-externalism, it is just the problem that came up with the causal/historical theories in the form of the "Madagascar" or "Santa Claus" examples.[316] Just as we want to say that what we now mean is dependent upon our past use, but leave room for the possibility that we occasionally break free from it, we want to say that what we mean can depend upon our future use, but leave room for the possibility of its breaking loose from us.

Objection: If a community were to split, with each sub-community developing accessible but incompatible equilibria from the original community's use, then it would seem that we would have to attribute both of the incompatible meanings to the original community's utterances.

Reply: Our entitlement to read our own development back into past use comes from our predecessors and ourselves being part of a single, developing, linguistic practice. Because of this, one can account for such cases by insisting that linguistic practices, like much else, do not preserve their identity through splits. The fact that, when such splits occur, the existence of a rival community undermines each community's entitlement to say what their predecessors meant does not entail that such entitlement is not present when no splitting occurs. An analogous point is frequently made with respect to personal identity: the fact that, if I were to split, I could not be identified with either of my successors does not, in itself, entail that I cannot be identified with my 'future self' when I don't split. An *actual* split undermines any identity between past and future, but the mere *possibility* of a split need not do so.

---

[316] See Evans 1973, Kripke 1972.

Objection: T-externalism still seems to entail at least that, when such splits occur, nothing determinate is originally meant by the term in question.

Reply: This objection is something we should take in stride. There are, after all, structurally similar cases within Burge's synchronic social picture and causal/historical accounts of proper names. That a community could split is no more a problem for the T-externalist than the possibility that a speaker's use of a name might be historically connected to two people is a problem for Kripke,[317] or the possibility that a speaker might unknowingly be part of two linguistically divergent sub-communities is a problem for Burge.[318] These possibilities are not *counterexamples* to the accounts in question; rather, they just suggest that, in some special cases, the accounts entail that nothing determinate will be meant by a term. This would be a problem if our intuitions suggested that something determinate was meant in these cases, but our intuitions about these cases tend to agree with the predictions of the accounts.

The intuition that we must *always* have thoughts with a determinate content must be given up if one is sympathetic with the causal/historical or social theories. The possibility of a split future thus does not present a novel type of problem that the T-externalist must face -- it is an instance of a phenomenon that all externalist (perhaps even all extensional) accounts must admit. Indeed, just as it should be viewed as a problem for a theory if it *did* come up with a determinate referent in the "Napoleon" case, it is a virtue of the above account that it predicts that there will be no determinate extensions in such split future cases.

---

[317] See, for instance, Evans' discussion (Evans 1973) of the possibility of our information about "Napoleon" actually coming from two different sources (one of whom did everything up until the Italian Campaign, and another of whom fought at Waterloo).

[318] Stich tells a story of this sort relating to the British and American use of "endive" and "chicory." The extensions of the two terms are reversed in the two countries, and so a speaker who spent an equal amount of time in each community, and felt the pull of both sets of norms, would fail to mean anything determinate by either term. (Stich 1983, 63.)

Objection: Another worry relates to the possibility of the use of the term having no future at all. If the Druids died out before the planes came by, the meaning of "ave" would remain indeterminate between *flying thing* and *bird*, since there would be no future linguistic behavior to make the distinction between them. The possibility of the term's having no future use need not, of course, be so apocalyptic. The term might simply fall out of use and be replaced by another, or undergo a change in meaning. (If their term "ave" gradually came to stand for just pillows, then their reactions to the first planes would no longer be relevant.) As a result, T-externalism potentially leaves many of our utterances without any determinate content.

Reply: In such a case, one should (once again) simply admit that the meaning of the term in question is indeterminate. After all, the suggestion that what Edwin meant by "ave" remains indeterminate if the "ave"-using practice died out before planes were encountered should match our pre-theoretical intuitions about such cases. Furthermore, since all the alternatives to T-externalism leave what Edwin's term refers to indeterminate in these cases as well, this can hardly be viewed as a reason for rejecting T-externalism in favor of one of these alternatives.

In any case, the increased visibility of this sort of indeterminacy cannot be held against T-externalism itself. Adding the community's future linguistic development can only *add* to the determinateness of the contents in question, so any scenario where contents remain indeterminate for the T-externalist will also be indeterminate according to more temporally bound versions of externalism. The T-externalist framework may make this type of underdetermination more *visible* but it can hardly be blamed for the *existence* of such underdetermination.

Objection: T-externalism seems to tie our meaning anything determinate to our community's having a non-branching future, but the determinacy of what we mean frequently seems independent of the future of our linguistic practice. For instance, by

1750, our ancestors would have meant *gold* by 'gold' even if they had all died out before they acquired the discriminating knowledge needed to pick it out as accurately as we do.

Reply: Of course it may be the case that the meaning of many, if not most, of our terms is independent of how things turn out in the future. However, this only shows that, for these terms, there is only one accessible equilibrium. As explained in section 4.1, a term's extension can be "settled" at a moment if the term has the same extension in every history passing through that moment. The meanings of many of our terms (those for which there is a single accessible equilibrium) may be settled in this fashion. A term with an unsettled extension, however, should not be understood as having an 'indeterminate' extension, since having an indeterminate extension is itself something which can be settled. According to the T-externalist, a term with an unsettled extension could still have a determinate extension in *every* history, while a term with an indeterminate extension will not have a determinate extension in *any* history.

Our use of "gold" in 1750 may have been such a case of a term's extension being settled, while the Druids' use of "ave" manifestly is not. By 1750, the practice in which our use of terms like "gold" was embedded was explicit and developed enough to make a phenomenological or functional interpretation of "gold" inaccessible. The Druids could not, by hypothesis, have such a well-developed normative structure in place governing their use of "ave."

It seems, then, that many of the reasons for thinking that there could not be a T-externalistic equilibrium for our semantic terms are not, in the last analysis, very compelling. The ultimate accessibility of any T-externalistic equilibrium will, then, depend upon the comparative accessibility of possible equilibria that take the contents of our utterances always to be settled at the moments of their utterance. Such accounts are the subject of the next section.

7 Non-temporalist interpretations of the phenomena

Any speaker's or community's usage up to a given time can underdetermine precisely what their terms refer to. Because of this, if one is to reject T-externalism and insist that all the relevant facts about an utterance's content must be in play by the moment of utterance, then one must claim that this underdetermination corresponds to an actual indeterminacy in what is meant. A non-T-externalistic semantics must thus be one that is tailored for indeterminate expressions. An obvious candidate of this sort is Field's partial reference based account of the semantics of languages with just such expressions.[319] Furthermore, while Field's account will be the focus of discussion in this chapter, but many of the criticisms that will be made of it (that it doesn't distinguish underdetermination from ontological confusion, that it fails to preserve continuity of meaning, etc.) can be carried over to other non-temporally based accounts such as those explaining such expressions in terms of degrees of truth. (None of the criticisms of Field's account have to do with his claim that certain utterances are neither true nor false.) How plausible we find T-externalism will depend to a large extent on how workable we find such alternatives.

While both the Druid and the Zebra cases are initially discussed by Wilson, Wilson himself suggests that the Druids' initial use of "ave" should be analyzed in terms of something like Field's notion of partial reference. According to the partial reference theorist, if, at a given time, the Druids' language (call it *D*) equally supports two rival interpretations, *I*1 & *I*2, that treat their term "ave" as referring to birds and flying things respectively, then "ave" *partially denotes* birds, and *partially denotes* flying things. Since the term partially denotes more than one (type of) thing, it doesn't *fully denote* anything. (A term fully denotes something if that thing is the only thing it partially denotes.)[320] The interpretations *I*1 & *I*2 both *partially accord* with the semantics of *D*, since they both assign

---

[319] Field 1973, 1974.
[320] Field 1974, 210.

to the words in *D* entities that the terms in *D* partially denote. Sentences in *D* can then be understood as true (false) if they are true (false) relative to *every* interpretation that partially accords with *D*. Sentences in *D* that have different truth values relative to different interpretations that partially accord with *D* are neither true nor false. On such an account, Edwin's sentence "Some aves are tasty" would be true, since it would be true on both I_1 & I_2. On the other hand, his sentence "Aves are found only on our island" would be false, since it is false on both interpretations. Finally, his sentence "All aves lay eggs" would lack a truth value, since it is true on I_1, and false on I_2.

Field uses his partial reference apparatus to explain Newton's use of the word "mass" (with "mass" partially denoting rest mass and partially denoting relativistic mass), and the T-externalist need not argue with this particular characterization. Indeed, this is a major difference between T-externalism and its more temporally bound competitors. The T-externalist can admit that there are some cases where, say, a partial reference account will be correct, but the partial reference theorist cannot allow cases where the T-externalistic semantic story is correct.

There is no denying that something like a partial reference account is appealing in some cases. For those cases where a partial reference account seems plausible, a T-externalistic semantics for cases of indeterminacy can be given that is structurally very similar to that proposed by Field.[321] In particular, one could give a partial reference semantics for those cases where a term's meaning is 'settled indeterminate.' A term's meaning will be settled indeterminate in precisely those cases where partial reference accounts will be most appealing. In particular, when the practice of using a word splits into two sub-practices (as was the case with "mass"), when the practice dies off, when no equilibrium is available for the entire practice, etc. The sentences that come out as true or false on Field's model would have the same values on such T-externalistic theories, and for those sentences that involve words whose extension is settled indeterminate, the T-externalist semantics could also be

---

[321] Or, if one prefers accounts that rely on degrees of truth to Field's account, one can build such mechanisms into one's T-externalistic semantics as well.

identical to the partial reference one. It is only *some* of the sentences that the partial reference theorists take to have no truth-value that the T-externalist need consider to be settled in the future.

On the other hand, while T-externalistic semantics can incorporate most of the mechanisms associated with Field's account, the converse is not the case. *All* of those terms whose extension the T-externalist can treat as "unsettled" the partial reference theorist must treat as having an indeterminate extension. (Compare the claim that statements about the future are neither true nor false to the claim that their truth value is unsettled.) Because of this, while various examples can be given where partial reference accounts can seem very plausible, the T-externalist can absorb such examples into his account. The debate must focus, then, on precisely those cases that the T-externalist appeals to, since it is only on these that there will be a difference between the respective judgments of the theorists in question. The question becomes, then, whether the class of terms whose extension is underdetermined by current use should be seen uniformly as having indeterminate extensions (as the partial reference theorist suggests), or whether they should be divided into two classes, settled indeterminate and unsettled (as the T-externalist suggests).[322]

Indeterminacy has almost always been presented as a semantic version of the underdetermination of theory by evidence, and underdetermination is typically presented as the possibility that two rival theories could both be compatible with all the available data, with the added claim that, in the case of semantics, there can be no 'fact of the matter' beyond that which can be determined. For most of what follows, the phenomena in question will simply be discussed in terms of underdetermination, since whether underdetermination at a time is actually indeterminacy is precisely with at issue between the T-externalist and someone who thinks that the facts available at the time are all the relevant facts.

---

[322] While Field originally presents the notion as one having to do with scientific terms and scientific revolutions, not terms in general, he seems to imply that his solution applies to all cases of indeterminacy.

How one feels about treating all cases of underdetermination as cases of indeterminacy will depend, to a certain extent, upon how important one takes the differences between the possible types of underdetermination to be. Consider, for instance, the following three possible types of underdetermination facing an interpreter.[323]

1. Underdetermination in which the extension of one of the rival concepts completely encompasses that of the other(s).

2. Underdetermination in which the rival concepts, while not coextensive, coincide (largely or) completely in the speaker's environment.

3. Underdetermination in which there is no overlap at all between the possible referents.

Underdetermination of the first sort can be seen in the initial use of "Grant's Zebra," and the second sort can be seen in Edwin's initial use of "ave." Cases of the third sort, however, are completely different. One is faced not with a common source of information that can be conceptualized in two different ways, but, rather, two distinct sources of information. This sort of underdetermination arises when, say, a speaker uses the term "Peter" for what he takes to be a single person, when 'Peter' is actually a pair of twin brothers, each of whom encounters the speaker equally often. Cases of the first two sorts should thus be distinguished from cases of the third, where the interpretee is 'ontologically confused.'[324] An interpretee who is ontologically confused is very different from one whose use merely underdetermines what he means by his terms. While semantic underdetermination should be understood as occurring when two or more rival semantic theories can both be compatible with all the available semantic evidence, cases of ontological confusion clearly aren't instances of this. It's not that *both* theories (i.e.:

---

[323] This list is not meant to be exhaustive. In particular, the underdetermination discussed here could be referred to as "practical underdetermination." There is no disagreement about what particular things the speakers are responding to, only about how those things should be classified. There is also, of course, "Quinean" underdetermination, which claims that the question of which particular objects the speakers are responding to will be underdetermined by the observational evidence (how can we tell they aren't responding to temporal slices of bird/flying-things, etc.). This latter purported type of underdetermination will not be discussed here (though for a further discussion of it, see Jackman, forthcoming).

[324] For the use of this term to describe cases of type 3, see Camp 1987, 21.

"Peter" refers to twin #1 and "Peter" refers to twin #2)[325] fit all the available semantic evidence. Rather it's that *neither* does. Picking between the two twins would be like choosing between two rival physical theories each of which can only explain half of the phenomena. The evidence does not, admittedly, allow us to choose between the two, but this is because it determines that neither alternative is empirically adequate.

There are thus serious disanalogies between cases of ontological confusion and cases of underdetermination in physical theory. As a result, such cases are not best understood as being instances of semantic underdetermination. In terms of the apparatus from chapter one, what distinguishes semantic underdetermination from ontological confusion is that in the former case there are *multiple* equilibria available to make sense of the interpretee, while in the later there are *no* such equilibria available. Characterized in this fashion, Newton's use of "mass" is a candidate for being understood as a case of ontological confusion. If Newton's belief that momentum was the product of mass and velocity, and his belief that the mass of an object did not change with respect to its frame of reference, were both deeply enough entrenched, there would be no single equilibrium available for his use of "mass," and some form of bifurcation would have been necessary. As a result, while rest and relativistic mass 'overlap' at low speeds in a way similar to how birds and flying things overlap on Edwin's island, Newton may best be understood as ontologically confused in a way that Edwin is not because he cannot give up either of his conflicting general beliefs in the casual fashion Edwin can give up his belief that, say, all aves are living things.[326] As a result, even if something like partial reference is the best way to deal with Field's example, one should not assume that it is a the best way to deal with underdetermination in general. Indeed, it would be preferable to give underdetermination a semantic treatment that is

---

[325] And these certainly aren't the only two available. One could also imaging theories which cashed out the confusion in their account of the speaker's predicates rather than in the satisfaction clauses associated with his singular terms.

[326] Ultimately, not much hangs on the correctness of the characterization of Newton as ontologically confused. The fact that the subsequent practice split is enough to legitimate Field's ascriptions in the T-externalist's eyes.

distinct from ontological confusion. The T-externalist is able to do this, the partial reference theorist is not.

Furthermore, it should also be noted that partial-reference theories are not especially hospitable to self-knowledge either. Explaining the Wilson cases in terms of partial reference commits one to saying that we change the meaning of our term each time the extension is made more precise and one of its partial denotations is eliminated. This is a very unattractive feature of partial reference accounts, since (as stressed in chapter three) a theory of meaning should, if possible, be phenomenologically faithful to our assumption that the meanings of our terms stay fairly constant over time.[327] Since the speakers often do not perceive themselves as changing what they mean by their words, a theory that says that they do must claim that they lack self-knowledge to at least that extent.[328]

Finally, it should be noted that all the considerations mentioned earlier in favor of the T-externalist account (increased determinacy, preservation of meaning over time, etc.) count against the partial reference one. These considerations are not, of course, decisive reasons for rejecting partial reference based accounts, but making room for a T-externalistic equilibrium need not involve showing that there are no alternatives available. Rather, it only requires showing that the T-externalistic equilibrium is palatable and that its competitors are not *obviously* preferable. Since the cases in dispute arise when a presupposition about language breaks down, it may be that any candidate equilibrium will

---

[327] The appeal of this intuition is not lost on Wilson, who suggests that meaning be held constant during a term's denotational refinement. Wilson states that while the extension of a term can change when we learn more about the world, there is no reason to think that the meaning has thereby changed as well, and he justifies this by claiming that:

> From the druids' point of view, the meaning of 'bird' has not altered, nor is there reason for the field linguist to describe matters in this fashion either. Attributions of change of meaning seem best tied to *recognition* by the community of a need for a conventional decision, and *ex hypothesi*, this is not the case here. (Wilson 1982, 572.)

[328] Though, once again, there must be room for this to occur in cases of prolonged mistaken identity turning into change of meaning.

pinch in some places. As was discussed in the first chapter, we have no guarantee that there will be a single right answer about what equilibrium to reach in such disputes. Just as Linnaeus and Ishmael can disagree about whether or not to call whales are "fish," without either party being guilty of any cognitive failing. The T-externalist and partial reference theorist could disagree over what to say about Edwin without either of them being mistaken. The purpose of this chapter has not been to show that one *must* adopt a T-externalistic equilibrium, but rather to show that if there are (multiple) equilibria available for our semantic terms, the T-externalistic equilibrium will be one of them.

8. Conclusion

It was suggested in the first chapter that disputes about meaning and content could be understood as the product of our trying to find an equilibrium between the general beliefs we associate with our semantic terms and the ascriptions we make when we attribute thoughts and sayings to others. The second and third chapter stressed the importance of our self-interpretations when reaching such an equilibrium, and argued that our practice of interpreting others in terms of social usage could be incorporated into an equilibrium because speakers understand *themselves* as engaging in a shared practice when using language. In much the same way, this chapter has argued that we can endorse everyday ascriptions of content that reflect a sensitivity to future usage because we understand this shared practice as itself extending through time. Most of the potentially unintuitive consequences of allowing future use to contribute to what we mean were dealt with on one of two ways. First of all, by distinguishing possible from accessible equilibria in the way suggested in chapter one, many of the consequences that might have been associated with radical shifts or splits in usage disappear. Secondly, many of the consequences (such as, say, those associated with self-knowledge and behavioral explanation) that come from adopting a T-externalistic position have already been accepted in virtue of incorporating the socially infected ascriptions Burge appeals to. Furthermore, the ways in which those

consequences were made palatable for the non-individualist discussed in chapter three work equally well for the T-externalist. Alternative ways of describing the Wilson cases were discussed and found to have a considerable number of unintuitive consequences of their own. Finally, it was argued that since attempts to find a more temporally-bound equilibrium for such cases turn out to be no more intuitive than their T-externalistic competitors, there is no reason why we cannot endorse our temporally sensitive ascriptions in these cases.

Summary and Conclusion

Finally, we should briefly summarize the argumentative thread in order to illustrate how the first three chapters, while each addressing a separate issue, support the T-externalistic position outlined and defended in chapter four.

The primary aim of the first chapter is to give an adequate account of semantic norms in terms of the requirement that we find an "equilibrium" for the putative extensions and general characterizations of our terms. As the discussion of Linnaeus's and Ishmael's use of "fish" illustrated, accepted usage at a time may leave a number of different equilibria open to the speakers of a community, and the differing interests and commitments of different speakers can lead them to adopt different equilibria and thus come to mean different things by their terms. This result is of considerable importance on its own, and is particularly relevant for the T-externalist once we come to understand disputes about meaning and content in terms of our trying to find equilibria for our semantic terms. In particular, the possibility that *multiple* equilibria may be accessible from our present usage makes room for a type of pluralism when it comes to such disputes. Their may be a type of 'blameless disagreement' between competing semantic theorists. As a result, the T-externalist need not tackle the ambitious task of showing that we are rationally *compelled* to adopt his position and that the competing accounts are unacceptable. Rather, he need only take on the more modest task of showing that we are at least rationally *permitted* to accept his position and the competing accounts are not obviously preferable to his.

The second chapter is concerned with defending and providing an account of the expressive constraint. The argument for the expressive constraint supports theT-externalist in two ways. First of all, by tying belief and utterance content together, the expressive constraint insures that a T-externalistic equilibrium for either thought or utterance content will be an equilibrium for the other as well. Secondly, the more synthetic account of the constraint presented at the end of the chapter introduces the idea that a speaker's conception of language itself plays a crucial role in determining the language he speaks and thus the

content of his thoughts. The T-externalist makes crucial use of this idea when he argues that our conception of language is, in fact, a conception of a shared practice that extends through time.

The importance of our self-interpretation plays a crucial role in the argumentative strategy of the third chapter as well. While the third chapter's manifest objective is to argue that an equilibrium for our semantic terms should endorse the truth of the sorts of non-individualistic ascriptions brought to our attention by Burge, it has a secondary agenda of paving the way for the T-externalistic position defended in chapter four. In particular, many of T-externalism's purportedly counterintuitive consequences come from its denying that one's thought contents supervenes upon non-relational facts about one's body. Non-individualistic accounts of content also require giving up the supervenience claim. As a result, if the non-individualist can give an acceptable account of self-knowledge and behavioral explanation that does not require such supervenience, such an account will be available also to the T-externalist. Most of the potentially counterintuitive consequences of allowing future use to contribute to what we mean (particularly those associated with self-knowledge and behavioral explanation) have thus already been defused once we have succeeded in incorporating the non-individualistic ascriptions Burge appeals to.

The fourth chapter tries to give a clear account of the T-externalistic position itself. Such an account makes use of the first chapter's explanation of when an equilibrium's is "accessible" in order to restrict the instances in which future usage can properly be understood as contributing to what we presently mean. Such clarifications, along with the dialectical moves established in the third chapter, are enough to disarm most of the objections to T-externalism. Since non-temporalist accounts of what to say about the cases that the T-externalist focuses on have problems of their own, a T-externalistic equilibrium seems accessible for our semantic terms. The problems with the non-temporalist accounts are not, however, devastating. One may be ultimately be able to adopt such an equilibrium for one's semantic terms as well. This would entail that our semantic terms lend

themselves to the type of pluralistic position the possibility of which was allowed for in the first chapter. The T-externalist and his more temporally conservative competitor may be like Linnaeus and Ishmael, each with an acceptable way to give a consistent account of most of the pre-systematized subject matter, but neither able to convince the other that their own systematization is preferable.

APPENDICES

Appendix A: Causal theories of reference -- normativity without general characterizations

It has occasionally been suggested that Kripke and Putnam's 'causal theory' of reference shows how the world's essential structure can help us fund a distinction between a term's putative and actual extension without having to resort to the mediation of the general characterizations associated with a word (thus allowing a term's meaning to be determined by its putative extension without being *identified* with it). On such an account, a term's putative extension at the time of its introduction is enough to determine its reference, since what the term refers to can be understood as everything that has the same 'essential structure' as most of the initial sample. A natural kind term picks out the class of things that has the same 'essential' physical structure as most of the items that fall under the term's original putative extension.[329] Proper names, presumably, have the entire person as the initial sample with the extension of the name being determined by the person's essential properties. (That these essential properties are unlikely to map on to any sort of natural kind associated with the biological sciences obviously presents a problem if one is hoping to find a *physicalistic* account of the reference of these terms).

Such an account would have little, if anything, to say about terms that were not natural kind terms, but one could argue that the robust response-independent notion of objectivity we find with purported natural kind terms such as "gold" will not be found with 'social kind terms' such as "chair."

Nevertheless, appealing to causation to give an account of meaning independent of general characterizations is not a very promising project. In particular, for the initial sample 'baptized' to name a person or natural kind, there would seem to need to be some sort of sortal implicit in the baptism (I am not introducing a proper name as a name for the person's family or jaw, and most samples of one natural kind are also instances of others).

---

[329] For such an account see Devitt 1980 (esp. pp. x, 27-31, 138).

Each baptism makes an at least implicit use of some sortals, and these sortals are reflected in the baptizer's general beliefs about the type of thing that he is baptizing.

It should also be noted that even leaving aside problems relating to the claim that the world has such an 'essential structure,' these purely causal accounts are not especially plausible as readings of either Kripke or Putnam.[330] First of all, the causal reading only really works for natural kind terms (and possibly proper names of people), while it seems fairly clear that the models of language suggested by Kripke and Putnam have wider application. Admittedly, the presupposition that the items picked out in the baptism form a unified 'kind' is stressed most by Kripke with respect to natural kinds,[331] but there is nothing in the story presented in *Naming and Necessity* that forces one to say that the only acceptable kinds are natural (rather than functional, legal or culinary) ones. Kripke states that, when the initial "baptism" takes place, "the object may be named by a description,"[332] and while including demonstratives in the description may pave the way for the more indexically driven accounts that purely causal theorists favor, that there be *any* such demonstratives does not seem at all essential to Kripke's picture. (See his discussion of "Neptune.")[333] Because of this, the reference of terms such as "contract" or "sofa" could be fixed by a description, and speakers unaware of the content of this original description could still refer to whatever the description picked out in virtue of being part of the chain of speakers who used the word. Indeed, Kripke's discussion of the term "meter" seems to presuppose that this is the case. A similar extension to non-natural kind cases can be found in Putnam's proposals about the 'hidden indexicality' in our language.[334] Admittedly, Putnam will talk about "hidden structures,"[335] and tends to focus on natural kind terms

---

[330] A claim defended recently, and with semantic issues in mind, by David Lewis. (See his discussion of 'elite properties' in Lewis 1983.)

[331] See, for instance, Kripke 1972, 120-1, 134-6, 138.

[332] Kripke 1972, 96.

[333] Kripke 1972, 79.

[334] Putnam 1975, 234, 271.

[335] Putnam 1975, 241.

like "gold,"[336] but he goes on to claim not only that "'pencil' is just as indexical as 'water' or 'gold,'"[337] but also that his conclusions "apply to the great majority of all nouns, and to other parts of speech as well."[338]

Furthermore, it seems clear that, even in the case of natural kind terms, the reference of the terms in question do not correspond to the natural kind instantiated by most of the items that fall within the putative extension because the way science carves up the world is 'metaphysically privileged' in a way that allows the world's 'essential' structure to impose itself on our words. Rather, it does so because the type of structures discovered by science correspond to the deeply entrenched presuppositions we have about the types of things we are talking about when we use kind terms. That is to say, it is our general characterizations that help determine which parts of the putative extension are 'really' correct. As a result, if we had a different set of general characterizations, different types of kinds would be picked out by exactly the same mechanisms that picked out the natural ones. For instance, Putnam claims that the 'hidden structures' determine the reference of natural kind terms, not because only such hidden structures could enter into causal relations with the baptizer, but rather because "normally the 'important' properties of a liquid or a solid, etc., are the ones that are *structurally* important."[339] The hidden structure of the initial sample determines the extension of the term because hidden structures are what are important to the baptizer and the speakers of the language. This view has as an obvious consequence that if other types of features are taken to be more important by the baptizer (and, as Putnam stresses, "importance is an interest relative notion"),[340] other types of features would be treated as 'essential' and thus determine how the initial set of examples is to be extended to a kind.

This difference of interest can be found even among those who agree that the term should be treated as a natural kind term. Consider, for instance, Putnam's discussion of

---

[336] Putnam 1975, 235.
[337] Putnam 1975, 243.
[338] Putnam 1975, 242.
[339] Putnam 1975, 239.
[340] Putnam 1975, 239.

how a molecular biologist and an evolutionary biologist might disagree over what was 'essential' to being a dog. A 'dog' that was found to be molecularly and genetically indistinguishable from 'real' dogs, but was known to have a completely separate evolutionary history, would be considered a dog by the molecular biologist's conception of species identity, but not by the evolutionary biologist.[341] Non-natural kind terms can thus all be accounted for in the baptismal picture simply by allowing that a baptizer can have different interests, and thus consider different types of properties 'important,' when they introduce new terms. These different interests can, of course, be understood in terms of (possibly implicit) general characterizations associated with the terms, and so if interests are brought in, one ends up with a dynamic picture.

That something like general characterizations are in the background of Kripke's picture is also clear when he says that "even though we don't *know* the internal structure of tigers, we suppose… that tigers form a certain species or natural kind."[342] It is this supposition that determines what parts of the putative extension do and do not fall under the term's actual extension. A similar thought is found in his suggestion that everyone could be wrong about whales being fish because responsibility to the internal structure of the animals in question was "part of the original enterprise" of classification.[343] Furthermore, even if one was willing to admit that scientific taxonomies were somehow 'more real' than non-scientific ones, this itself would do nothing to show that our everyday terminology 'really' corresponds to some scientific taxonomy.

Kripke's and Putnam's responses to the overemphasis on general characterizations found in theories like the descriptive theories of names is thus best understood as providing

---

[341] Putnam 1993, 77.

[342] Kripke 1972, 120-1. Furthermore, both Kripke and Putnam are willing to say that these general characterizations which structure our practice can turn out to be mistaken. Kripke is happy to admit that "we may be mistaken in supposing that there is such a kind" (121), and when we have made such a mistake, we would not go on to say that there are no tigers. The same line of thought can be seen in Putnam's suggestion that, in spite of the fact that we assume that they are animals, we could imagine discovering that all of them are robots.

[343] Kripke 1972, 138.

a more dynamic picture of what was involved, not as advocating a theory of meaning where the putative extensions (with the help of the world's 'essential' structure) are overemphasized in a corresponding fashion. Given that they were reacting to one extreme (meaning determined solely by general characterizations), it is not surprising that their own work can be easily (mis)understood as advocating a recoil to the corresponding extreme (general characterizations play no role at all), but such a reading not only makes their views seem fairly implausible, but also reduces their significance by missing the extent to which the stories about natural kinds and proper names can be understood as particular examples of a much more general phenomenon.

Appendix B: Semantic defects as epistemological virtue

Accounts that tie meaning too closely to putative extensions while ignoring the general characterizations are unable to give any substantive account of the possibility of our being wrong about the correctness of our application to those items in (the core of) the putative extensions. However, there has been a tradition of trying to make a virtue out of this defect in one's semantic theory. In particular, if one's account cannot make sense of the possibility of one's being mistaken in a certain range of cases, then the account leaves no room for skeptical doubts about these cases. As a result, rather than being criticized for underfunding our conception of objectivity, such accounts of meaning have on occasion actually been praised for their purported 'anti-skeptical' consequences.

Tying what a term means to its putative extension[344] and then appealing to what the term means to fend off skepticism is characteristic of so-called 'paradigm case' arguments. Paradigm cases tend to be the most deeply entrenched members of a term's putative extension, so if what a term means is solely a function of its putative extension, there will be no way for these central applications to turn out to be mistaken. An instance of this strategy can be seen in Stebbing's response[345] to Eddington's claim[346] that desks were not 'really' solid.

> What are we to understand by "solidity"? Unless we do understand it we cannot understand what the denial of solidity to the plank amounts to. But we can understand "solidity" only if we can truly say that the plank is solid. For "solid" is just the word we use to describe a certain respect in which a plank of wood resembles a block of marble, a piece of paper, and a cricket ball, and in which each of these differs from a sponge, from the interior of a soap-bubble, and from the holes in a net. [347]

---

[344] Or at least the 'core' of the putative extension: these would be the parts of the putative extension that are the most deeply 'entrenched,' perhaps the applications made in certain 'optimal' conditions (for something like "that is red") or items that are 'prototypical.' Of course, such deeply entrenched items will be much less likely to be given up than others, but being wrong in such cases is not impossible. If they are given up, then perhaps the rest of the putative extension will be given up too (i.e.: nothing will be taken to satisfy the term in question). Still, while one will expect this to be a fairly rare occurrence, it can still happen (even 'paradigm cases' of witches will turn out not to be, since there are, of course, no witches).

[345] Which is endorsed by Urmson as a successful application of "the argument from standard examples." (Urmson 1953, 121-2.)

[346] In Eddington 1929.

[347] Stebbing 1937, 51-2.

Even if one does agree that physics has not shown that tables are not solid, the argument from paradigm cases is far from conclusive. It relies on the meaning of terms like "solid" being determined solely by their putative extensions, and so leaves itself open to responses such as the following:

> The effectiveness of Eddington's remark seems to derive from the fact that the properties which one is inclined to regard as the defining properties of solidity do not really belong to the objects to which we customarily apply the term 'solid,' and the interest of this fact is in no way diminished by repeating 'Oh, but we *do* apply the word "solid" to such things as desks.' Of course we do, but we also say that by 'solid' we mean 'having its interior entirely filled with matter,' or something like that. The collision is *within* ordinary usage, and not between it and scientific theory.[348]

Eddington took physics to have shown that most of what fell under the putative extension of 'solid' did not fit the general characterization we associated with that term. While Eddington may have given too much to the general characterization of "solid," Stebbing focused too much on the term's putative extension. The question of whether tables are *really* solid or not will inevitably involve comparing both.

And if the two cannot be reconciled, (i.e.: there is no real equilibrium involving both) one might want to say that there is, in fact, no equilibrium for our previous use of "solid," and that the term needs to be replaced by two homonyms, one tied to the putative extension, another to the general characterization. Both the claim that science has shown that tables aren't 'really' solid and the claim that it has shown that solid objects 'really' consist mainly of empty space presuppose that an equilibrium can be found that dumps most of either the general characterizations or the putative extension.

Much recent work in epistemology informed by the 'externalist' semantic tradition also seems to focus too much on the putative extensions of our terms, particularly with reference to the central and 'paradigmatic' examples.[349] Externalism about content

---

[348] Mates 1958, 127. Mates distinguishes two approaches to discovering what a person means by a given term. The "extensional" approach (which focuses just on the putative extension the speaker associates with the term) and the "intensional" approach (which attempts to get the speaker to come up with general characterizations). I am fairly sympathetic not only with this distinction, and but also with his suggestion that the results of these two methods need not coincide. (See Mates 1958, 125-6.)

[349] Similar epistemic work can be done, of course, by overemphasizing the aspects of the general characterizations. Take, for instance, the following example from C.I. Lewis:

encourages people to think that systematic perceptual error is impossible because the contents of our perceptual terms are determined precisely by those objects we apply those perceptual terms to.[350]   Nevertheless, while it may be true that widespread *perceptual* error could be ruled out, this certainly does not mean that the perceptually-based judgments must be true. There have been many witch sightings, and, since there are no witches, we can assume that all these perceptually based witch-beliefs were false. What is at issue here is not 'perceptual error,' rather it is that our general characterization of "witch" was not satisfied by any of the people who were called "witches." That the meanings of our terms could be affected by these general characterizations makes the transition from the impossibility of general perceptual error to anti-skepticism about perceptual judgments problematic.[351]

Consider, for instance, a speaker who refers, as we do, to both the morning star and the evening star as "Venus." When Venus rises at the beginning of the evening he may say "there is Venus." When he is asked what he is looking at, he will reply "Venus." He does the same in the morning, and never 'misapplies' the term to any other star or planet. The putative extension of "Venus" in his language consists only of the planet Venus. However, while the putative extension of Venus in his language is much like it is in ours (assuming we can identify Venus in the night sky), the general characterization he associates with the term is very different. In particular, while we believe that Venus is a lifeless planet, that we have sent satellites to explore it, etc., this speaker believes that Venus is the goddess of love, that she is married to Vulcan, is the lover of Mars, responds to the prayers of her devoted followers (of whom the speaker is one), etc.

It is an a priori principle that physical things have mass. By this criterion they are distinguished from mirror-images and illusion.… the truth of the principle is independent of the particular phenomenon. But a world in which we should experience phenomena having persistence and independence not characteristic of imagination, and a coherence not characteristic of our dreams, but things which would still not be amenable to any gravitational generalizations, is entirely conceivable. In such a world our a priori principle would not be rendered false -- since it is definitive of the physical; but the category "physical" might well be useless. (Lewis 1929, 26-7.)

[350] See, for instance, Burge 1986b, 131.

[351] This is, I take it, the gist of Nagel's complaint about Putnam's Brain-in-the-vat argument. (Nagel 1986, 71-3 .)

Now it should be fairly clear that for his term "Venus," there is a serious lack of fit between the general characterizations and what he actually applies the term to. If the speaker is a particularly devout worshiper, he may be much more willing to give up everything in the putative extension rather than his general beliefs. If he finds out that 'Venus' is a lifeless planet, he will be inclined, correctly, to think that he misapplied the term every morning and evening when he looked at the sky.[352]

The holistic nature of the dynamic between our terms' putative extensions and general characterizations may insure that, if our terms pick out anything at all, the most deeply entrenched elements of our usage will turn out to be true, but there is no guarantee that anything from a particular term's putative extension will be among the most deeply entrenched set.[353]

---

[352] There may be some temptation to treat utterances of his such as "Venus is looking especially bright tonight" as still being true (provided that the planet is especially bright on that night), but this is because it is true in the sense that my utterance of "John looks hurt" might be considered true if I mistake Peter for John and Peter looks hurt. The speaker's reference for such utterances may be Peter or the planet, but the semantic reference will still be John or the Greco-Roman goddess.

[353] Nor, for that matter, will there be any guarantee that anything from the general characterization will be among such a set.

Appendix C: On "The Methodology of Naturalistic Semantics"

In his recent paper,[354] Devitt points out that "we start semantics in the unusual position of having to specify a subject matter," and while he thinks that "we should not insist on great precision about this in advance of theory," he still insists that "we do need some explication of our vague talk of 'meanings' … that is *not ad hoc*."[355] He conceives of this as just one instance of a more general type of problem:

> What constitutes some property we ascribe, "being an F"? The question may often have little scientific interest, or course, but where it does -- presumably "being a gene," "an atom," "an acid," "an echidna," and "a pain" are examples -- *how do we tell*? Sometimes we already have a well-established theory; for example, thanks to molecular genetics, we have for "being a gene." But suppose that we do not and are starting pretty much from scratch, which is surely the right supposition for semantics. How then do we tell what is common and peculiar to Fs "in all possible worlds?"   What is the "ultimate" method?[356]

Devitt describes this "ultimate" method as follows:

> It breaks into two stages. First we must identify some apparently uncontroversial examples of F's and non-F's. Second, we must examine the examples to determine the nature of "being an F." …This second stage is a straightforwardly scientific one. It is hard to accomplish and, as is well known, even harder to say how we accomplish it .… We should note, however, an important feature of this second stage: the examination can lead us to reject some of the results of the first stage; apparently uncontroversial examples turn out to be controversial. Thus we may conclude that some of the things identified as Fs are not; for example, whales are not fish. We may conclude that some things identified as non-Fs are Fs; for example, tomatoes are fruits. We may even conclude that *none* of the things identified as Fs is an F; for example, there are no witches. Usually, such revisions in our preliminary identifications take place only if our examination leads to a powerful theory.[357]

While the contents of this chapter commit me to treating something like this proposal of Devitt's as right, Devitt's own characterization of the 'ultimate method' has a number of serious flaws that prejudice his account of semantics (or perhaps simply reflect antecedent prejudices about semantics).

As the title of his paper suggests, Devitt wants his account of semantics to be "naturalistic," and is unapologetic about wanting to "bring semantic methodology close to other scientific methodologies," since, after all, "according to naturalism, semantics is an

[354] Devitt 1994.
[355] Devitt 1994, 548.
[356] Devitt 1994, 561-2.
[357] Devitt 1994, 562.

empirical science like any other."[358] This naturalistic methodology is supposed to, among other things, constrain our reliance on intuitions in semantics:

> Clearly , there could be no objection to the standard appeal to intuition in semantics if the semantic task were simply the systematization or explanations of the intuitions. Some write as if they think that this is the task. Yet this task should be no more interesting or appropriate than that of systematizing or explaining our ordinary physical, biological, or psychological intuitions. The intuitions in each case are *about* the subject matter that should concern the science, they are not the subject matter themselves.[359]

> In the face of the dominant role played by intuitions in semantics, it is worth laboring the point. From the naturalistic perspective, semantic intuitions are like those in any other science: open to revision in the face of empirical theory. We could be wrong about what has putative meaning. We could be wrong in thinking that anything has it. Even if we are right that something has it, we could be wrong in thinking that it plays a semantic role and so really is meaning.[360]

Intuitions associated with most of the 'standard' thought experiments thus play a role in the first stage of the "ultimate method," since these attitude ascriptions will be the source of the "uncontroversial examples" to be gathered during the first stage.[361]

Devitt is right to think that our semantic intuitions associated with various thought experiments should be understood as playing something like this role, and his complaints about the *ad hocness* of various semantic proposals that do not follow anything like the ultimate method seem justified.   Nevertheless, his own insistence that semantics be naturalistic seems to be an instance of precisely the type of tenacious devotion to an *a priori* claim about content that an ultimate method should free one from. His version of the ultimate method cannot do this since it is only supposed to work for natural kind terms.[362] This becomes especially clear when we consider that the second stage is supposed to be "a straightforwardly scientific one."[363] As a result, the method would only work for natural

---

[358]   Devitt 1994, 545.

[359] Devitt 1994, 546. See also, "the 'basic' task is to say what meanings *are*, to explain their *natures*." (Devitt 1994, 547.)

[360] Devitt 1994, 568. "Putative meanings" are discovered by the "descriptive task" of giving an account of how we use our semantic terminology. Deciding if these putative meanings play the semantic roles of explaining behavior and providing truths about the world (Devitt 1994, 550) (and thus if they are *real* meanings) is taken to be part of the "normative task" (Devitt 1994, 565).

[361] Devitt 1994, 566.

[362]   "Gene," "atom," "acid," "echidna," "pain," are all, presumably, taken to be such terms.

[363] Devitt 1994, 562.

kinds, and thus, if it is to work for semantic terms at all, semantics would have to be naturalistic.

However, since the assumption that semantic terms should be understood naturalistically is undermotivated, the fact that Devitt's version of the ultimate method requires this assumption suggests that Devitt's characterization of the ultimate method is not general enough.[364] Indeed, we can find a more general account of how we find "what constitutes some property we ascribe, 'being an F'" in Tyler Burge's "Intellectual Norms and the Foundations of Mind,"[365] with Burge's "examples" and "archetypical applications"[366] corresponding to Devitt's first stage and his "normative characterizations" corresponding to something like Devitt's second stage. For a natural kind term, the task of coming up with a normative characterization based on the examples may, of course, be a "straightforwardly scientific one," but Burge makes it clear that the method is intended to apply equally well to such non-natural kinds as "chair," "baby," "walk," or "knife."[367]

In addition (and related) to its lack of generality, Devitt's account of the ultimate method suffers from treating the dialectic between the uncontroversial examples and general characterizations too simply. In particular, it treats the dialectic as if the general characterizations came only after scientifically investigating (or at least reflecting upon) the uncontroversial examples. This seems unrealistic; for most terms we will have both a set of uncontroversial examples of and uncontroversial beliefs about the items in question, and any ultimate method will involve finding an equilibrium between these two. Take, for instance, Devitt's own example of "witch." It does not seem possible that we came to the conclusion that there were no witches by anything like the method Devitt proposes. In particular, it seems highly unlikely that we first identified a number of 'uncontroversial

---

[364] Devitt seems, at times, quite happy to admit the arbitrariness of this assumption (Devitt 1994, 558).

[365] Burge 1986c.

[366] Burge 1986c, 703.

[367] And, of course, Rawls can be understood as applying something like the ultimate method for ethical terms like "justice" ("reflective equilibrium" being the ideal point where the products of the two stages correspond (Rawls 1971)). Indeed, this general 'Socratic' procedure has a fairly long and distinguished history.

examples of witches' and then through scientifically investigating these examples came up with a characterization of what witches are that none of these examples satisfied. Furthermore, even if there were a property that all and only the purported witches had (red hair, warts on the nose, non-matching eyes, whatever), we would still have concluded that there were no witches, not that "witch" turned out to be synonymous with, say, "redhead." We concluded that there were no witches because we have a number of strong general beliefs about witches (they can cast spells, work for the devil, etc.) that we are not inclined to give up even if none of the "uncontroversial examples" of witches turn out to satisfy them.

There can be a number of such general beliefs that are brought into an investigation rather than derived from it, and the assumption that a term picks out a natural kind may be just such an assumption.[368] Furthermore, just as reflection and general theories about what it is to be an F can cause us to reject some of the originally uncontroversial examples, they can also cause us to reject some originally uncontroversial assumptions. The assumption that a term picks out a natural kind is just such an assumption, and there are many terms (for instance, "tree," "lily," "reptile," and "fish") for which this assumption has turned out to be false.

When such presuppositions conflict with the results of investigation, we may have a number of options. Consider the case of the word "ape." This term was applied uncontroversially to orangs, chimps and gorillas. It was also not applied to humans (we were descended from apes, but we were not apes ourselves). The term was also assumed to pick out a natural kind. Unfortunately, there is no biological kind that includes orangs, chimps and gorillas and does not include humans as well.[369] Given this fact, we are faced with three options: (1) admitting that we are, in fact, apes (treating "apes" being a natural

---

[368] That the 'real' semantic properties explain behavior and that they tell us about the world are also presupposed by the investigation rather than derived from it, though one would expect that these general beliefs are a good deal less negotiable than Devitt's naturalism.

[369] Gould 1983, 358.

kind as non-negotiable, our not being apes as negotiable), (2) admitting that "apes" does not, after all, pick out a natural kind (treating "apes" being a natural kind as negotiable, our not being apes as non-negotiable), or (3) admitting that there are no apes (treating both as non-negotiable). By building naturalism into the ultimate method itself, Devitt would force us to take either the first or the third option, even though the second is, in many ways, the most appealing.

If the idea that semantic terms pick out natural kinds (i.e.: that semantics can be "naturalistic") is just one general idea that can be brought into the dialectic associated with any real ultimate method, then it obviously should not be a precondition of any ultimate method being used at all. Though one could, of course, take the idea that semantic terms pick out natural kinds to be non-negotiable in the way in which witches' casting spells may have been,[370] in which case, it could simply turn out that there were no meanings, beliefs, etc.[371] This option, which could push one towards a form of eliminative materialism, does not seem especially desirable, and while it is just an unappealing option on the more general method described above, it seems as if it could very well turn out to be a necessity on Devitt's. A more general methodology would make 'folk psychological' terms like "belief"

---

[370] Though reasons for thinking that we should not treat naturalism as being non-negotiable are given by Devitt himself:

> Suppose that our actual practice of ascribing certain properties serves our purposes in some area well. That is very good evidence that we ought to ascribe at least those properties to serve those purposes. The situation in semantics appears to be of this sort. For consider how successful our ordinary attitude ascriptions seem to be at serving their purposes. Their apparent success in explaining behavior is certainly not limitless but it is nonetheless impressive. Their apparent success in influencing our own thoughts about the world could hardly be exaggerated. So it is likely that, at least, we ought to ascribe to tokens that are thought and uttered the properties that we do ascribe and hence that those properties are meanings. Evidence for a descriptive theory will be evidence for a normative one and hence the basic one .… In the absence of a detailed comparison with the semantic status quo, a revisionist proposal will have a credibility problem: it will always seem much more plausible that there is a flaw in the philosophical arguments offered in support of the proposal than that the apparently successful status quo is so radically mistaken. (Devitt 1994, 561)

> The assumption that an account must be naturalistic may be just the type of flaw that exists in arguments attempting to overthrow the status quo. After all, the success of a practice certainly doesn't imply that the terms of the practice will necessarily map on to natural kinds (functional kinds being just one obvious alternative around which successful practices could be based).

[371] An outcome that Devitt seems to consider extremely unlikely, though still possible (Devitt 1994, 561).

no more threatened by the possibility of a failure of naturalistic reduction than 'folk biological' terms like "tree" or "lily."

Appendix D: Methodological and Ascriptional Individualism.

The purpose of this appendix is to distinguish "methodological" individualism from "ascriptional" individualism. In the body of the chapter, the term "individualism" is used to pick out ascriptional individualism, and while this sort of individualism is rejected, methodological individualism is not criticized. It will be argued in this appendix that writers such as Davidson and Bilgrami occasionally defend ascriptional individualism because they take their commitment to methodological individualism to entail it.

Methodological individualism requires that factors 'external' to the interpretee (such as, say, the usage of the surrounding community or the physical structure of what is around him) be counted as relevant in determining what he means only if such factors are taken to be relevant by the interpretee (in its crudest form, he takes "arthritis" to mean "what the experts mean by 'arthritis,'" or he takes "water" to mean "whatever has the same atomic structure as most of the stuff I've called 'water'" etc.). Methodological individualism is thus incompatible with a strong anti-individualistic position according to which we can be assured, *a priori*, that a speaker's words should be interpreted according to the community standard regardless of the speaker's attitude towards such a standard. In Davidson's terms, methodological individualism requires that all linguistic norms must ultimately be grounded in the intentions of the agent himself.[372] Davidson puts forth a methodologically individualist position in the following passage.

> If the speech behavior of others does not provide the norm for the speaker, what can? The answer is the intention of the speaker to be interpreted in a certain way provides the norm; the speaker falls short of this intention if he fails to speak in such a way as to be understood as he intended. Under usual circumstances a speaker knows he is most apt to be understood if he speaks as his listeners would, and so intends to speak as he thinks they would. He will then fail in one of his intentions if he does not

[372] A charitable reading of *Naming and Necessity* is thus compatible with methodological individualism. While Kripke is, justly, unsympathetic with metalinguistic attempts to give a sense to proper names, as stated earlier, it is an *essential* part of his story that "the speaker intends to use the name in the same way as it was transmitted to him" (i.e., he intends to be a 'reference preserving link' in the chain of communication). There are problems with Kripke's initial formulation of the idea (see appendix F), but it is clear that, for Kripke, the relevant norms only get a grip on the speaker though the mediation of his intentions. This was, of course, an aspect of Kripke's account that Devitt tried to get rid of in giving his purely 'causal' account. As a result Kripke's position, unlike Devitt's, could still be understood as methodologically individualistic. (Devitt's cannot because he suggests that the 'real essences' of the kinds in our environment affect what we mean independently of any concern we have with them.)

speak as others do. This simple fact helps explain, I think, why many philosophers have tied the meaning of a speaker's utterance to what others mean by the same words … (where "others" refers to a linguistic community, "experts," or an élite of one sort or another). On my account, this tie is neither essential nor direct; it comes into play only when the speaker intends to be interpreted as (certain) others would be. When this intention is absent, the correct understanding of a speaker is unaffected by usage beyond the intended reach of his voice.[373]

This claim about the indirectness of the connection is quite plausible. While we are usually willing to consider ourselves subject to the norms of the surrounding community, if we decide, in certain cases, deliberately to flout communal usage it is hard to see why what we mean should still be tied to communal usage. If I insist on calling stools "chairs" in spite of having the fact that my use is non-standard pointed out to me on many occasions,[374] and I make clear to people unfamiliar with my usage how it differs from the norm,[375] then it seems perverse to describe my calling a stool a "chair" as the expression of a false belief. It would be better to say that I just meant something different by "chair" than the rest of my community.

Still, one can agree with Davidson that if one lacks the intention to be interpreted as others would be, then one shouldn't be so interpreted, and still argue that in most cases, we do intend to be interpreted as others would be.[376] Claiming that the norm to which I am

---

[373] Davidson 1992, 261.

[374] I reply in these cases "Well, I just mean something different by "chair" than most people do."

[375] Davidson stresses how important it is that I make sufficient effort to make my idiosyncratic usage known.

> If a speaker reasonably believes he will be interpreted in a certain way, and speaks with the intention of being so understood, we may choose to say he means what (in the primary sense) he would have meant if he had been understood as he expected and intended. Reasonable belief is such a flexible concept that we may want to add that there must be people who would understand the speaker as he intends, and the speaker reasonably believes he is speaking to such a person.… These remarks should make plain why Dummett's accusation, that I endorse a variety of Humpty Dumpty's theory that meaning depends only on intention, does not find its target. (Davidson 1994a, 12-13.)

> Still, if one really must be able reasonably to expect others to be able to interpret one correctly, one might think that departures from the social norm could only be exceptions made intelligible against a background of general conformity, and that one could not reasonably expect to be properly understood if one departed from the surrounding norms for *all* (or even most) of one's words. Especially if, like Davidson, one thinks that disagreement is only intelligible against a shared background of agreement.

[376] This need not, of course be an explicitly formulated prior intention, but can be seen as part of a background to our linguistic intentions (see appendix F), which leads us to understand ourselves as mistaken when we defer to accepted usage. (Indeed, deference itself may be seen as involving this intention as a type of 'intention in action.' Provided, of course, one thinks the concept of an intention in action makes sense -- which Davidson clearly does.)

beholden in using a word is determined by how I intend to use the word[377] is not incompatible with non-individualism if how I intend to use the word usually turns out to be tied to the linguistic practices of the surrounding community.[378] In particular, we can assume that Bert intended to be interpreted as the doctor would be, and his failed intention to speak correctly did foil "the intention to be interpreted a certain way." Indeed, the fact that Bert is speaking to a *doctor* and not, say, his brother is relevant here. When we are actually speaking to experts in a context in which their expertise is being appealed to, the presumption that we will intend to be following 'expert usage' will be considerably stronger than when we are talking within a possibly idiosyncratic sub-community such as our family. There is little reason to think that someone couldn't accept Davidson's point and still insist that *Bert* should be interpreted the way Burge does.[379] It may be the case that, when Bert complains to his brother, "I hate this cold weather, it always makes my arthritis worse," he should be interpreted as meaning *tharthritis* by "arthritis," but the case Burge describes is not like this at all. Bert is going to see a *doctor* and in this context (consulting a doctor) it seems quite plausible to think that Bert intends to be following expert usage, his utterance of "I have arthritis in my thigh" having (very roughly) the gist of "I know there is this disease you doctors call "arthritis" and I think that I have *it* in my thigh." Such a 'metalinguistic' reworking of his communicative intention has little plausibility in the case of his complaint to his brother, and Burge's choice of context has a lot to do with the plausibility of his example. After all, if he *didn't* believe that the doctor knew more about arthritis than he did, it isn't clear why he would be going in for a

---

[377] Intentions "give content to an attribution of error by allowing for the possibility of a discrepancy between intention and accomplishment." (Davidson 1992, 259.) The appeal to community is, of course, not the only way to generate cleavage between how I intend to use a word and how I actually use it. Any criteria by which I can 'take back' various attributions of the term as 'mistaken' will be enough to generate the difference.

[378] Of course Davidson claims that "a failed intention to speak 'correctly,' unless it foils the intention to be interpreted in a certain way, does not matter to what the speaker means" (Davidson 1992, 261), but the non-individualist can respond that "our intention to be interpreted in a certain way" is often directly connected to our intention to speak 'correctly.'

[379] After all, Burge himself claims that "the individual's intentions or attitudes toward communal conventions and communal conceptions seem more important than the causal antecedents of his transactions with a word" (Burge 1979, 114).

consultation in the first place. The tie between the meaning of a speaker's utterance and what others mean by the same words may be neither essential nor direct, but it is often there nonetheless.

Methodological individualism is thus compatible with our occasionally ascribing contents in a fashion that this chapter characterizes as "non-individualistic." We can say, for instance, that Bert means *arthritis* by "arthritis" if he intends to use the term to pick out what the doctors do by it. Davidson endorses such a possibility in passages such as the following:

> I do not doubt the existence of [the linguistic division of labor], or even its importance.…we can take it to be part of the meaning of an expression that its reference is to be determined by expert opinion.… So for the words "elm" and "beech" to pick out the appropriate trees there would have to be experts, but we cannot conclude that the meaningful use of these words demands a social setting.… It is obvious that the linguistic division of labor can't be essential to verbal communication. If everyone meant by "elm" "what others mean by elm," the word would have no reference. The linguistic division of labor is a device that can come into play only after the basic linguistic skills that tie words directly to things are already in place. So no matter how universal the linguistic division of labor is in practice, it cannot constitute the essential social element in language. We could get along without it.[380]

Davidson here admits that, for *most* of us, what we refer to by terms like "elm" may be determined by experts,[381] so if the experts had referred to something different, we would have too. These 'non-individualistic' content ascriptions are, however, justified on purely *individualistic* grounds. 'Accepted' usage is relevant only because the *individual* takes it to be. Davidson seems only to be denying that, for any given word, this can be true of *all* of us, and thus that the division of linguistic labor is essential to language as such. His real target is, then, not the claim that the truth conditions of our thoughts and utterances may be in part determined by the usage of the surrounding community, but rather the stronger claim that such dependence is a *necessary* feature of language use.[382]

---

[380] Davidson 1994a, 5-6. In much the same way, Bilgrami seems to endorse the claim the we can ascribe contents non-individualistically if the speaker himself intends to be interpreted that way. (See Bilgrami 1992a 23, 42.)[381] Davidson, of course, contests this very point in a number of places, including the pages immediately following the passage quoted above.

[382] And this claim about language and communication *as such*, need not be in conflict with Burge's claim that the fact that we "normally look to others to help set standards for determining the range of legitimate examples and the sort of background information used in explicating a word or concept .… derives from psychological necessities for human beings." (Burge 1989, 186, 187).

These certainly aren't the only places where Davidson seems to commit himself to a position that is ascriptionally non-individualistic. Davidson not only claims that "a social theory of interpretation" is made possible by belief taking up "the slack between sentences held true by individuals and sentences held true (or false) by public standards,"[383] but also argues that "states of mind like doubts, wishes, beliefs, and desires are identified in part by the social and historical context in which they are acquired."[384] Furthermore, ascriptional individualism does not sit well with Davidson's discussion of "porcupine" in the following passage:

> I may not know the difference between an echidna and a porcupine; as a result I may call all the echidnas I come across porcupines. Yet, because of the environment in which I learned the word 'porcupine,' my word 'porcupine' refers to porcupines and not echidnas; this is what I think it refers to, and what I believe I see before me when I honestly affirm, "That's a porcupine."[385]

It is not at all clear what the relevant structural difference is supposed to be between Davidson's porcupine/porcupine-or-echidna example and Burge's arthritis/tharthritis example.

Indeed, Davidson here seems to commit himself to there being cases where what we are talking about should be understood determined by our social environment. Davidson's reference to "the environment in which I learned the word" should not be understood as a reference merely to his *physical* environment, since Davidson cannot be suggesting that he refers to porcupines because he learnt the word "porcupine" when he was in contact with members of the natural kind *porcupine*.[386] Not only has he expressed skepticism about the idea of natural kinds,[387] but he also explicitly insists that the phenomenon is not limited to natural kind terms and is "ubiquitous, since it is inseparable from the social character of language."[388]

---

[383] Davidson 1974a, 153. He further claims that belief is a private attitude that "is not intelligible except as an adjustment to the public norm provided by language" (Davidson 1975, 170).
[384] Davidson 1989, 170.
[385] Davidson 1989, 168.
[386] Which, for most of us, would be an implausible claim about how the term was learnt.
[387] Davidson 1987, 450.
[388] Davidson 1989, 167.

Ascriptional individualism leaves no room for the intermediate position that methodological individualism allows for. In particular, if Bert believes that he has "arthritis" in his thigh and applies the term to a pain in his thigh, then the ascriptional individualist will claim that there is no way to understand him as having misapplied his term, and thus no way to understand him as referring to *arthritis* by "arthritis." On such an account, it is not merely the case that semantic (lexical) normativity must ultimately be grounded in the speaker's intentions, but also the case that there is no room for semantic norms to really go beyond perceptual errors. Ascriptional individualism is thus considerably stronger than methodological individualism, and it is incompatible with the type of ascriptionally non-individualistic position that the methodological individualism allows for.

Both Bilgrami and Davidson, however, occasionally advocate versions of ascriptional individualism in passages such as these:

> I am not impressed by [Dummett's] or Burge's or Putnam's insistence that words may have a meaning of which both speaker and hearer are ignorant. I don't doubt that we sometimes say this, and it's fairly clear what we have in mind: speaker and hearer are ignorant of what would be found in some dictionary, or how people with a better or different education or a higher income use the words. This is still meaning based on successful communication, but it imports into the theory of meaning an elitist norm by implying that people not in the right social swim don't really know what they mean.[389]

> The notion of a mistake is a normative notion, and the possibility of mistakes of this sort introduces a matter of rightness or wrongness regarding the use of terms or the application of concepts the terms express. Rightness or wrongness of this sort is essential to the idea of meaning and concepts, and therefore to the idea of content. My view does not allow one to say of Bert and others like him that they have made or can make these sorts of mistakes.… when it came to describing Bert's deference to his doctor, I did not describe it as learning from the doctor that he was all along mistaken about arthritis and had been wrongly applying his term for it, but rather as the doctor having helped him acquire a new concept so that he was now wielding a new term. Therefore my view has no place for an essential normativity that attaches to meaning and content.[390]

While Davidson and Bilgrami defend both ascriptional and methodological individualism, their motivation seems to be tied more to methodological individualism, and both endorse claims that are incompatible with ascriptional, but not methodological, individualism. Indeed, the considered positions of both of them are probably versions of methodological

---

[389] Davidson, 1994a, 6, 9. It should also be noted that, in those cases where we use a word idiosyncratically, we can count on it to "work" in the sense Davidson requires.

[390] Bilgrami 1992a, 80.

rather than ascriptional individualism.[391] If their defending ascriptional individualism is just an aberration,[392] why do they so often do so?

One reason would be that the two types of individualism are occasionally not distinguished clearly enough and that, in defending one, they feel that they must defend the other. This is especially clear with Davidson. While Davidson is certainly right in thinking that we can give up the intention to follow community norms in particular cases,[393] one suspects that he occasionally runs together such cases (where we deliberately flout the societal norms) with those where we are merely ignorant of them. In the former case we lack the intention to be interpreted as others are, while in the latter case such an intention[394] is probably present, and this difference is enough to make it the case that, even if the first type of speaker should be interpreted individualistically, the latter type of speaker usually should not.[395] While there are occasions where we are willing to be interpreted idiosyncratically if our usage departs from the social norm, the intention to be interpreted as others are usually is the default case. Davidson occasionally suggests that just the opposite is true: that a willingness to be interpreted idiosyncratically should be understood as the default. It should, however, be clear that, in most contexts the default assumption should be that we will be interpreted according to standard usage, and so unless we are willing to flag a willingness to depart from such usage, we should intend to be interpreted in such a fashion. This phenomenon may be even more pronounced in certain more 'disciplined'

---

[391]   Indeed, in correspondence, Bilgrami explicitly endorses methodological over ascriptional individualism: "Bert has his own criteria beliefs about arthritis in each locality he uses it. Sometimes he may have in mind to say it about something that causes hurting in his thigh in the mornings, sometimes he may use it to be talking about whatever the doctor talks about when he uses it etc." (Personal correspondence, July 1995)

[392] Though they both come out of a tradition that seems pretty clearly committed to ascriptional individualism. Quine's original discussion of "radical translation" in *Word and Object* very explicitly ruled out interactions with "native kibitzers" whether the informant was disposed to rely on them or not.

[393] These cases would presumably be limited to those words that refer to items for which we feel confident that we could set up adequate criteria without anyone's help (I may be able to do this with "chair" but not with, say, "electron," or, for that matter, "elm").

[394] Or at least the assumption/presupposition that we will be so interpreted.

[395] Excepting those cases discussed in section 1.211 of this chapter (slips, malapropisms, and radical misunderstanding).

fields where there is a concerted effort to see to it that there is a clearly defined accepted usage (technical vocabulary) and participants in the field are expected to adhere to the accepted usage when using such vocabulary. We generally intend to be interpreted as our fellows are, and while such an intention is defeasible (if, for instance, our usage strays too far from the public norm),[396] minor differences between ours and the public use will not be enough to make us give it up.

Davidson's concern about what is essential to language *as such* need not be relevant to what is, in fact, true for *our* language. As a result, the fact that language may turn out not to be *essentially* social need not affect the claim that contents *actually* ascribed to us should often be individuated non-individualistically. Davidson's interest in what is conceptually necessary for communication[397] often leads him to overlook the fact that what is at issue with cases like Bert's use of "arthritis" is not what language *must* be like for *anything*, or even what it *must* be like for *us*, but rather how it is, *in fact*, for *Bert*.

Furthermore, both Davidson and Bilgrami interpret Burge as defending a position that is incompatible with even *methodological* individualism,[398] and because of this, they naturally feel that Burge's arguments for his position need to be criticized. However, in arguing against Burge, both tend to contest the accuracy of Burge's claim about how Bert's utterance to his doctor should be interpreted, suggesting instead that Bert should here be understood as meaning something different than the doctor does by "arthritis." (In much the same way, they argue that Bert's deference should be explained not by his understanding himself as misapplying his term, but rather by his understanding himself as meaning something different by "arthritis" than his fellows and changing what he means to

---

[396] What counts as 'too far' may be context-relative and determined in part by the judgments of the speaker.

[397] Davidson is interested in "the *essential* nature of the skills" that are "*basic* to linguistic communication" and not "usual, though *contingent*" features of such communication. (Davidson 1982, 278, 280 [italics mine].)

[398] Whether they are correct in doing so will not be discussed here. (Though Burge 1989 makes their reading of Burge considerably less plausible than Burge 1979 would have on its own.)

allow him to communicate more effectively with them.)[399]   However, arguing against Burge in this way implicitly commits them to ascriptional individualism, since Bert's interaction with his doctor is fairly close to a best-case scenario for allowing methodological individualism to justify ascriptionally non-individualistic attributions. Because of this, if we must interpret *Bert* individualistically *here* it seems quite plausible that we will have to interpret *everyone* individualistically *everywhere*. As a result, their strategy for defending methodological individualism implicitly commits them to ascriptional individualism.

If one wants to defend methodological individualism, there is no need to quibble with Burge about whether or not Bert has made a false claim about arthritis. The methodological individualist can freely admit that Burge is right about Bert's utterance to his doctor. All they need to do is argue that (1) such ascriptions are compatible with methodological individualism since they can be explained by Bert's reliance on the doctor's use in that context, and that (2) in contexts where no such reliance is present, there is no reason to ascribe contents in a social-externalist fashion. A defender of methodological individualism need not challenge the *correctness* of Burge's characterization of Bert's utterance, he need only challenge its *generalizability*.

As a result, then, methodological individualism, which clearly is an essential part of the Davidsonian program, can accommodate the types of ascriptionally non-individualistic intuitions that Burge appeals to. Consequently, Davidson and Bilgrami need not understand Burge's work as the kind of threat they frequently treat it as being.

---

[399]   See sect 1.222 of this chapter for a more extended discussion of this. Methodological individualism leads to ascriptional individualism if we deny that the speaker could have (or attach much weight to) the sorts of beliefs that would build a reliance on others into his concepts (such as the belief that doctors know much more about arthritis than he does), and such an account of deference would go a long way towards justifying such a denial.

Appendix E: Ascriptions *de dicto* and *de re*.

The accounts of self-knowledge and behavioral explanation presented in this chapter rely heavily on the *de dicto/de re* distinction, and the existence of such a distinction should be fairly uncontroversial. Both those who see the distinction as corresponding to two types of belief,[400] and those who see it instead as corresponding to two types of ascription,[401] share the assumption that belief predicates are ambiguous.[402] However, there are many ways of drawing the *de dicto/de re* distinction, and the purpose of this appendix is to clarify the version of the distinction that is presupposed in this chapter.[403]

The reasons for thinking that belief predicates are ambiguous should be clear when we consider an ascription such as:

(I) Oedipus believes that his mother is the queen of Thebes.

There seems to be a sense in which the ascription is clearly true, and a sense in which is clearly false. It can seem true because the woman who was, in fact, Oedipus' mother was believed by him to be the queen of Thebes. It can seem false because Oedipus himself would have vigorously and sincerely dissented from the suggestion "your mother is the queen of Thebes." Belief ascriptions thus seem to admit of two interpretations,[404] and this would suggest belief predicates are ambiguous.

However, while this chapter relies on the *de dicto/de re* distinction at a number of crucial points, it will be argued below that the impression of ambiguity that one can get by focusing on cases such as (I) is just another example of how hard cases make bad law.

---

[400] See, for instance, Kaplan 1968, Burge 1977.

[401] See, for instance, Searle 1983 (esp. pp. 196, 217), Dennett 1982, Brandom 1994 (esp. ch. 8).

[402] As Stich recently put it "just about everyone who has written on [the *de dicto/de re* distinction] in the last twenty-five years has shared Quine's view that belief predicates are systematically ambiguous." (Stich 1986, 119.)

[403] It should be noted that the characterization of the distinction does not involve a focus on the types of 'epistemically strong' *de re* beliefs that are associated with perceptual judgments and demonstrative utterances.

[404] Once again, leaving it open whether these two interpretations should be taken to correspond to two types of belief or two types of ascriptions.

When we consider our ascriptional practices as a whole, the suggestion that belief predicates are ambiguous has little plausibility.

Why this is so can be seen when we look at Brandom's recent and detailed characterization of the *de dicto/de re* distinction as one between two types of ascription. Brandom characterizes these two types of ascriptions and their relation to the *de dicto/de re* distinction as one between "two different styles in which the content of the commitment ascribed can be specified."[405] Consider, for instance, the following two ascriptions of belief to Oedipus:

(I) Oedipus believes that his mother is the queen of Thebes.

(II) Oedipus believes that his mother is the queen of Corinth.

Both (I) and (II) are ascriptions that can truly be made, the first because Oedipus believes that Jocasta is queen of Thebes, and the second because he believes that Merope is queen of Corinth. Of course, while we would endorse the characterization of Jocasta as "Oedipus's mother," Oedipus would not, and while Oedipus would endorse the characterization of Merope as "Oedipus's mother," we would not. According to Brandom, this "social difference of deontic attitude turns out to be what determines whether the ascription is *de dicto* or *de re*."[406] Ascriptions such as (I) in which we characterize Jocasta in terms that we, but not Oedipus, would accept, are *de re*, and ascriptions such as (II), in which we characterize Merope in terms that Oedipus (though in this case not we), would accept, are *de dicto*. As Brandom puts it:

> The substitutional commitments that govern the expressions used to specify the content of the commitment ascribed can . . . be either attributed to the one to whom the doxastic commitment is ascribed, or undertaken by the one ascribing it … [Where the ascriber employs] only commitments the ascriptional target would acknowledge, the content specification is *de dicto*. Where the ascriber has employed substitutional commitments he himself, but perhaps not the target, endorses, the content specification is *de re*.[407]

The *de dicto/de re* distinction is thus explained in terms of whether or not the content is ascribed in terms that either the ascriber or ascribee would accept. Such an account makes

---

[405] Brandom 1994, 503.

[406] Brandom 1994, 503.

[407] Brandom 1994, 503, 507.

it trivially true that *de dicto* belief ascriptions will capture the ascribee's point of view, while letting the *de re* ascriptions capture what the ascriber takes this point of view to be a point of view on. Because of this, the fact that what we refer to may be determined by factors of which we are not aware does not threaten our knowledge of our beliefs as ascribed *de dicto*, or allow us to ascribe inconsistent beliefs to an agent in *de dicto* ascriptions such as "Oedipus believes that his mother is not his mother."

Still, while it seems promising to distinguish the *de dicto* from the *de re* in ascriptional terms, this *distinction* cannot be understood as a *dichotomy*. The distinction does not reflect an *ambiguity* in our belief predicates. To see why, we should first notice that, on such an account, many, if not most, ascriptions qualify equally as *de dicto* and *de re*. For instance, suppose I make the ascription:

(III) Oedipus believes that his daughter is good to him.

The characterization of Antigone as Oedipus's daughter is both one that I endorse and one that Oedipus would acknowledge. Ascriptions such as (III) would seem to involve both undertaking and attributing the commitment to the substitution in question, and thus might seem to count equally as *de re* and *de dicto*,[408] which is why (III), like most attitude ascriptions, is not comfortably classified as just one or the other.[409]

How the ascriptional account can explain why most actual ascriptions seem neither purely *de dicto* nor purely *de re* can be shown by making clear how the two types of substitutional commitments (those accepted by the ascriber, those accepted by the ascribee) divide the possible substitution instances of a term into four classes.

1. Those acceptable to both the ascriber and the ascribee.
2. Those acceptable to the ascriber but not the ascribee.
3. Those acceptable to the ascribee but not the ascriber.
4. Those acceptable to neither the ascriber nor the ascribee.

---

[408] Though, according to the passage from Brandom quoted above, they would count only as *de dicto*.

[409] This is a distinct advantage the ascriptional account may have over those that treat the *de re*/*de dicto* distinction as one between two types of belief, since such accounts suggest that the belief in question must be one or the other (unless, of course, one is willing to say that there are two beliefs involved). (Ascriptional accounts also seem to deal better with iterated ascriptions. (See Brandom 1994, 608-613).)

Ascriptions of the first sort seem to be neither purely *de re* nor purely *de dicto*.[410] Ascriptions of the second sort could be characterized as 'pure' or 'paradigmatic' *de re*, while those of the third sort could be characterized as 'pure' or 'paradigmatic' *de dicto*. Belief ascriptions involving substitutional commitments of the fourth sort seem to be neither *de re* nor *de dicto*, but while they may not be very popular, there can be such ascriptions. In particular, there might be a substitution accepted by the ascriber's audience (or some other salient group) but which the ascriber wants to make clear that neither he nor the ascribee accepts. Such a case might be an ascription such as "Chomsky believes of those [s]freedom fighters[s] that they are ruthless terrorists."[411]

While it seems clear that most of our ordinary ascriptions are of type (1), cases of misidentification involve ascriptions of types (2) and (3).[412] For instance, if Nicole shoots a cow thinking that it was a deer,[413] we may characterize her beliefs with sets of ascriptions such as:

> Nicole believes of a cow that it is a deer. She thinks that, if she bags it, the [s]deer[s] will provide her with venison for the winter. She hopes that the [s]deer[s] won't suffer much, and doesn't want to hit the [s]deer[s] in the head, because she wants to mount the head on her wall.

Since the belief ascriptions most discussed in the philosophical literature (Superman cases, Pierre's puzzle etc.) often gravitate to types (2) and (3), it can be easy to lose sight of the fact that most of our belief ascriptions involve characterizations of the topic that are accepted by *both* the ascriber and the ascribee.[414] Because of this, we should not claim, as

---

[410] Though, once again, Brandom's official line seems to count both cases 1 and 3 as *de dicto*. However, on his regimentation, a *de dicto* ascription only commits one to being a case of type 3.

[411] [s]Scare quotes[s] are here used to mark the fact that the ascriber does not accept the substitution in question, and they will be used more here than the "so called" locution for terminological convenience.

[412] Of course, not all ascriptions of type (2) and (3) (especially *de re*) need involve mistakes.

[413] See Brandom 1994, 522ff.

[414] Proper names help ensure that most ascriptions will be of type (1), since they provide a type of common currency that makes such ascriptions easier. I can ascribe to my brother the belief that Cervantes wrote *Don Quixote*, and, while this ascription may turn out to be false, it is highly unlikely that it will be because, while he believes of Cervantes that he wrote *Don Quixote*, he does not believe it of him as "Cervantes." There are, of course, people we have beliefs about without knowing their names (X believes of Ortcutt that he is a spy), and cases where people have more than one name, but, in most cases, the use of a proper name will insure that the ascription will be of type (1).

Brandom does, that belief ascriptions are ambiguous.[415]   If all ascriptions were either of types (2) or (3), such a claim would be uncontroversial: there would be just two types of ascriptions.[416]   However, ascriptions of type (1) are not as easily viewed as being in any way ambiguous; they can express *both* a doxastic commitment and a doxastic attitude.

Since most ascriptions of belief are of type (1), one might argue that types (2) and (3) are (in spite of being 'pure' *de re* and *de dicto*) actually degenerate forms of belief ascription. In our unregimented language, cases of type (1) are the default cases. Unless we specify otherwise,[417] it is assumed that both the ascriber and the ascribee accept the substitutions in question. This is why (if one is unfamiliar with the story) statements like (I) and (II) can be misleading: our default assumption will be that both the ascriber and the ascribee are willing to endorse the characterizations in question.[418] Such statements would not be so misleading if there were a mere ambiguity involved, since there would be no such default assumption.[419] As a result, everyday belief ascriptions should not be seen as ambiguous between *de dicto* and *de re* ascriptions, but rather, pure *de dicto* and *de re* ascriptions should be understood as special cases in which we explicitly disavow some of the commitments associated with the everyday ascriptions. In particular, with *de dicto* ascriptions, we disavow any commitment to the substitution in question, while with the *de re* ascriptions, we disavow any commitment to the ascribee's accepting the substitutions in

---

[415] "'Belief' is ambiguous in scorekeeping terms, referring sometimes to deontic status and sometimes to deontic attitude -- sometimes to doxastic commitment and sometimes to acknowledgment of such a commitment." (Brandom 1994, 260.)

[416] Brandom's "regimentation" of English is suited for such a two way ambiguity. Cases of type (2) appear in Brandom's regimented language as "S believes *of* t that F(*it*)," while those of type (3) appear as "S believes that F(t)," but there is nothing in the regimentation to correspond to types (1) or (4).

[417] Or the context makes it obvious, as in "Grant believed that the site of the 1996 summer Olympics had to be captured."

[418] We can thus bring the semantics of such terms comparatively close to our actual intuitions about their truth conditions. Both Brandom's (1994) attempt to make the *de dicto* ascriptions the default and, say, Salmon's (1986) attempt to characterize all such ascriptions as *de re* must rely to a much greater extent on 'pragmatic' considerations to explain our (apparently mistaken) intuitions about the truth conditions of these ascriptions.

[419] I.e., for paradigmatically ambiguous words such as "bank," even if one of the two meanings has a certain 'default' status in various contexts ("I keep my money in the bank" etc.) it isn't the case that the default status ever picks out *both* of the putative meanings. We don't naturally assume that the speaker has in mind a financial institution on the side of a river.

question. These disavowals can be explicitly flagged with the "believes of" locution in the case of *de re* ascriptions and the "so-called" locution or ˢscare-quotesˢ in the case of *de dicto.* The simple "X believes that P" locution is reserved for ascriptions of type (1), where the characterization is acceptable both to the ascribee and the ascriber.

In response to this, someone who wanted to keep the independence of our *de dicto* and *de re* ascriptions might argue that, rather than characterizing ascriptions of types (2) and (3) as explicit disavowals from the commitments involved in ascriptions of type (1), we can take ascriptions of type (1) simply as a *combination* of the commitments involved in ascriptions of types (2) and (3).[420] The 'pure' *de dicto* and *de re* ascriptions would be primitive, and 'mixed' ascriptions of type (1) would be taken to be what you get when you put the two pure ascriptions together. However, there are at least two reasons to think that we should treat cases of type (1) as more primitive than those of types (2) or (3). The first is that our everyday ascriptions treat cases of types (2) and (3) as "marked"[421] cases while treating cases of type (1) as the default case. The second is that treating cases of type (2) and (3) as primitive makes it difficult to account for the possibility of cases of type (4). If one starts with just ascription-types geared towards ascribing the two types of commitment, it is hard to see how one is going to be able to use them to get a case where neither type of commitment is ascribed. On the other hand, if one takes cases of type (1) as primitive and cases of types (2) and (3) as produced by mechanisms that strip cases of type (1) of their usual commitments, then it should seem quite simple how one could get cases of type (4) by applying both mechanisms. Indeed, it would actually be a problem for this type of account if cases of type (4) were not possible.

---

[420] This is the line Brandom suggests when he talks about how the cases in which ordinary that-clauses both characterize the speaker's conception and pick out the relevant object can be regimented as, say "Alfred believes *of* the man in the corner *as* the man in the corner that he is a spy." (Brandom 1994, 546.)

[421] "Believes-of" locutions and scare-quotes (corresponding to the colloquial "so-called") respectively.

To return to the four types of substitutional commitments that can be expressed in a belief ascription, the suggested notation for them would be:

1. Those acceptable to both the ascriber and the ascribee.   John believes that Peter is bald.
2. Those acceptable to the ascriber but not the ascribee.   John believes of Peter that he is bald.
3. Those acceptable to the ascribee but not the ascriber.   John believes that ^{S}Peter^S is bald.
4. Those acceptable to neither the ascriber nor the ascribee. John believes of ^{S}Peter^S that he is bald.

Since this regimentation allows one clearly to represent all four cases (while attaching the normal form to ascriptions of type (1) and registering the degenerate character of types (2) & (3)), it is more recognizably English than Brandom's (which attaches the normal form to the *de dicto*,[422] and which only allows for two types of ascriptions).[423]

Such an account, in which the 'pure' ascription forms, while acceptable, are in some sense degenerate, may suggest that belief predicates, rather than being strictly ambiguous are governed by something like a preference rule system.[424]   Examples of preference rule systems can be seen in the application of verbs such as *climb*. Even though climbing typically involves both (a) upward motion and (b) effortful grasping motions ("Bill climbed (up) the mountain"), neither the first nor the second condition need be satisfied ("Bill climbed down the mountain" ; "The snake climbed up the tree"), though at least one of the two must be (* "The snake climbed down the tree"). This behavior is not best explained by suggesting that verbs like *climb* are ambiguous. The standard presupposition when we come across the verb is that *both* of the conditions are satisfied. Belief ascriptions could

---

[422] Brandom's mode of regimentation suggests that most actual belief ascriptions are *de dicto*, since it reserves the 'unmarked' "believes-that" locution for *de dicto* ascriptions, while the account here suggests that they are neither purely *de dicto* nor *de re*. Once again, Brandom does characterize cases of type (1) as *de dicto*; since *de dicto* attributions only involve attributing a substitutional commitment, his regimentation suggests that in everyday cases we undertake no commitment to the substitution in question.

[423] *De re* ascriptions appear in Brandom's regimented language as "S believes *of* t that F(*it*)," while *de dicto* ascriptions appear as "S believes that F(t)." As a result, if the 'regimented' forms correspond only to the 'pure' *de dicto* and *de re* ascriptions, then there is nothing in the regimentation that allows one to specify a case as being of class (1) or (4), while if they pick out the 'mixed' ascriptions (i.e. both (1) and (3) count as *de dicto* and both (1) and (2) count as *de re*) not only is there still no way to pick out cases of type (4) at all, but there is no way to specify that one has cases such any of the other three types (i.e., (1) as opposed to (1) or (2)) in mind.

[424] For a discussion of such "preference rule systems," see Jackendoff 1990, 35-6. (See also Lakoff 1987.)

be viewed as working in much the same way. The healthiest and 'prototypical' belief ascriptions involve substitution instances accepted by both parties, though attributions can still be made with substitution instances accepted by just one of the parties.[425]

---

[425] This would suggest that cases of type (4) would not be acceptable unless explicitly marked by the grammar or context. So "Peter believes that the inventor of bifocals was the first postmaster" would be maximally acceptable if both Peter and the ascriber thought Franklin invented bifocals, moderately acceptable if just one of us did, and not acceptable at all (or at least no more acceptable than "the snake climbed down the tree") if neither of us did.

Appendix F: Internalism, Individualism and the Background

Given that this chapter attempts to account for and justify ascriptionally non-individualistic practices within a methodologically individualistic framework, it is worth examining Searle's similar attempt to incorporate the truth of 'externalist' ascriptions within his 'internalist' framework. By discussing some of the virtues and faults of Searle's account, I will hope I will make the position defended in this chapter clearer.

Rather than challenging the intuitions Kripke and Putnam appeal to, Searle proposes to accept such intuitions and argue that they can be integrated into an 'internalist' framework.[426] Indeed, rather than treating them as counterexamples to his position, Searle takes Kripke's, Putnam's and Donnellan's examples to illustrate how internalistic contents should be specified.[427] Searle characterizes his position as "Holistic Internalism," which he defines as follows:

> What is inside the head is entirely sufficient to determine the identity of each of our intentional states. The various "causal" and "contextual" conditions referred to by the externalists are entirely represented by the mind. However, the conditions of satisfaction of each intentional state are only fixed relative to the Network and the Background which is not part of each intentional state though, of course, the Network and the Background are nonetheless parts of the mind. (Searle 1991, 237.)

Searle claims that critics of internalism have missed this alternative because they "look only at what the agent might *say* and not to the *total Intentional content* he has in his head, and also because they neglect the role of the Network and the Background."[428] Searle's suggestion that various 'background' capacities would allow one to justify ascriptionally non-individualistic content ascriptions is potentially very fruitful. Unfortunately, it is not one that Searle fully exploits.

To see why, we need to consider some of what Searle says about the "Background." The Background is crucial for Searle because "in general, representations only function, they only have the conditions of satisfaction that they do, against this nonrepresentational

---

[426]   Searle 1983, 203.
[427]   Searle 1983, 204, 206, 244-6.
[428]   Searle 1983, 250.

background."[429]   This Background is "a bedrock of mental capacities that do not themselves consist in Intentional states (representations)"[430] but which includes "various skills, abilities, preintentional assumptions and presuppositions, stances, and nonrepresentational attitudes."[431]   Furthermore, Searle is quite explicit that not only should elements of the Background not be viewed as intentional states themselves, but also they should not be seen as *part* of the content of our intentional states.[432]   As a result, even if we believe that facts about what we mean must ultimately be explained in terms of facts about what is going on inside our heads, if some of these facts are part of the Background, then they need not be built into the Intentional content. They affect the Intentional content by helping to determine the conditions of satisfaction, but they do not enter into a specification of the Intentional content itself.[433]

---

[429]   Searle 1983, 21.

[430]   Searle 1983, 143.

[431]   Searle 1983, 151. The following quotations should give a rough feel for how Searle understands the Background:

> If we actually tried to carry out the task [of specifying each of the intentional states in the Network] we would soon find ourselves formulating a set of propositions which would look fishy if we added them to our list of beliefs in the Network; "fishy" because they are in a sense too fundamental to quality as *beliefs*, even as unconscious beliefs. Consider the following propositions: elections are held at or near the surface of the earth; the things people walk on are generally solid; people only vote when awake; objects offer resistance to touch and pressure. (Searle 1983, 142.)

> A man certainly *could* have the belief that tables offer resistance to touch, but … that isn't the correct way to describe the stance that I, for example, now take towards this table and other solid objects. For me, the hardness of tables manifests itself in the fact that I know how to sit at a table, I can write on a table, I put stacks of books on tables … as I do each of these things I do not in addition think unconsciously to myself, "it offers resistance to touch." (Searle 1983, 142.)

[432]   Searle 1983, 19, Searle 1991, 230, 232-3.

[433]   As a result, one need not accept Searle's claim that "Even if I am a brain in a vat … nonetheless, I do have the Intentional content that I have, and thus I necessarily have the same Background that I would have if I were not a brain in a vat and had that particular Intentional content." (Searle 1983, 154.) One reaches such a conclusion only if one builds the Background into the Intentional contents themselves, rather than simply allowing that the Background helps determine what the Intentional contents are. I would have different Intentional contents if I were a brain in a vat, because the Background which the brain and I share produces different Intentional contents in the different contexts. (Incidentally, Searle's claim that I would have the same Intentional contents even if I were a brain in a vat points to a serious weakness in his discussion of the twin cases. Namely, he accounts for the fact that I and my twin-earth counterpart have different Intentional contents by the fact that we are different people and thus have different thought/experience tokenings. His account is thus not able to distinguish Intentional contents if we think of the *same* experience as arising in different counterfactual situations (as his discussion of the brain in the vat makes abundantly clear). As a result, just as Searle says that I would have the very same Intentional contents were I a brain in a vat, he must also say that I would have the very same

Unfortunately, Searle himself often fails to "distinguish those features of content which function to determine conditions of satisfaction from features outside the content which also serve to fix conditions of satisfaction."[434] In particular, Searle's specification of the conditions of satisfaction of our intentional states in causally self-referential terms seems to build aspects of the Background into the Intentional content itself.[435] As a result, in spite of his insistence that the Background is "non-representational,"[436] he goes on to claim that "the various 'causal' and 'contextual' conditions refered to by externalists are entirely *represented* by the mind."[437] Searle thus seems to miss his own point that while "it is always possible to take an element of the Background and treat is as a representation … from the fact that it is possible to treat an element of the Background as a representation, it does not follow that, when it is functioning, it is functioning as a representation."[438]

For instance, it is plausibly part of the Background that we understand ourselves as perceiving objects in an independent world that causally impinges upon our senses.[439] This understanding of our place in, and relation to, the world we perceive is manifested in our everyday behavior and can be understood as part of the preintentional Background. Because of this, it is quite plausible to think that the "causally self-referential" character of visual perception is a function of the Background, but Searle builds it directly into the perceptual content itself, claiming that the intentional content of, say, my visual experience of seeing a yellow station wagon before me should be spelt out as:

Intentional contents if I were brought up on Twin Earth instead of Earth, in spite of the fact that the people etc. I interact with are completely different.)

[434] Searle 1991, 232.

[435] Searle admits that he occasionally makes such an assimilation, but he is clear that such assimilations are only a type of "shorthand": "When arguing in favor of "internalism" . . . I sometimes use the expression "Intentional content" as a shorthand for Intentional content including Background and Network conditions on its conditions of application, all of which are internal to the mind." (Searle 1991, 233.)

[436] Searle 1983, 21.

[437] Searle 1983, 237, italics mine.

[438] Searle 1983, 158.

[439] This might be related to the commitment to 'realism' that Searle takes to be part of the Background. (Searle 1983, 159.)

> I have a visual experience (that there is a yellow station wagon there and that there is a yellow station wagon there is causing this visual experience).[440]

Searle's specification of the Intentional content thus ignores a possibility that his account clearly allows for: the content of the visual experience involves the yellow station wagon itself. Our perception has this content only thanks to elements in the Background, but these background elements are not part of the content. It need not follow from the fact that one must be able ultimately to give a first-personal account of our semantic concepts that our Intentional contents themselves should be specified in such an explicitly first personal manner.

While Searle's actual account of Intentional contents seems unsatisfactory, his general approach of stressing the importance of the background implicit in our practices can be extended in a number of productive ways. For instance, the Background can include not only a general picture of how we are causally related to the objects of perception, but also our 'implicit'[441] understanding of the 'informational network' we are part of. That is to say, while we may not explicitly think that, say, our memories refer to people who are causally responsible for those very memories, such an 'assumption' is implicit in our practices, and forms part of the Background against which our thoughts get the contents they have. If we remember something, we assume we were there, and if we discover that the person we take a memory of ours to be a memory of could not have been causally responsible for the memory, then we assume that we must have been thinking of someone else. As a result, a subject can be said to 'know which' object he is thinking of even if he cannot explicitly *state* the object's role in the informational system.

With these background capacities in place, one can satisfy Russell's Principle ("a subject cannot make a judgment about something unless he knows which object his

---

[440] Searle 1983, 48. Such an account of the content of our perceptual experiences is not meant to correspond to anything running through our mind when have such experiences, rather, it represents "a verbal specification of what the Intentional content requires if it is to be satisfied." (Searle 1983, 49.)

[441] Such talk of our understanding of something as 'implicit' should in no way be understood as suggesting that what is understood as 'represented' in any way. Rather, it only suggests that we will act in a way that will manifest a commitment to the truth of what is implicitly understood.

judgment is about")[442] in cases where Evans might not have originally thought it possible. For instance, Evans gives the following as an example of someone who could not satisfy Russell's principle, and thus not have the thought about a particular shiny ball.

> Our subject briefly sees one ball rotating by itself on one day, and the other on a later day. And let us further suppose that the subject retains no memory of the first episode, because of a localized amnesia produced by a blow to the head. Suppose, finally, that many years later our subject reminisces about 'that shiny ball' he saw many years earlier. If asked which ball he is thinking about, our subject cannot produce any facts which would discriminate between the two.[443]

Evans admits, of course, that the subject could specify the ball in question if he reflected on the role of memory in the informational system, but the resulting thought would be of an entirely different kind.

> It is true that *if S reflects on the process of information-transmission*, he may, so to speak, aim a thought (a particular-thought) at 'that ball.' But the object of the thought will be individuated, in this case, *by reference to its role in the operations of the information system*. This is a different kind of thought from the kind of particular-thought which it is the aim of this chapter to discuss.[444]

Furthermore, he claims that "it is surely correct to assume that thinkers (among whom we have to include very young children) will not in general resort to this [causal] way of identifying the object of their thought."[445] However, while the typical subject will not *appeal* to the object's role in the informational system, there is still a sense in which the subject *knows* that the object must play such a role, and can thus be said to know which object he is thinking of. Particular-thoughts may be of a different kind from those that make *explicit* reference to the object's role in the informational system, but there is no reason not to think that such thoughts may still rely on an *implicit* awareness of such a role to get their content.[446] When background capacities are brought into play, Russell's

---

[442] Evans 1982, 89.

[443] Evans 1982, 90.

[444] Evans 1982, 128. (The particular example has been changed from 'that native' to 'that ball' for continuity.) See also, "It is an assumption of the case that the subject would not have any notion of appealing to these causal considerations; otherwise the case would cease to be a counterexample to Russell's Principle" (Evans 1982, 117).

[445] Evans 1982, 117. See also: "I assume, as I think that I am entitled to, that he would not think of distinguishing the ball he is thinking of as the one from which his current memory derives." (Evans 1982, 117.)

[446] Evans is surely right to think that thoughts that bring these background capacities *explicitly* into play are of a different type from those that do not, and if this is the case, then one has another reason for thinking that, *pace* Searle, characterizations of the Intentional contents of our thoughts relying on such a background should not make explicit reference to their 'self-referential' character.

principle seems much less restrictive. We can allow that *merely* being causally connected to an object is not enough to have thoughts about it (which is the gist of the "photograph model" Evans sees as in opposition to Russell's principle), while also insisting that our implicit understanding[447] of the informational system helps constitute our ability to think of those items from which our memories derive.

The sorts of 'intuitive' considerations Kripke and Donnellan muster in support of their accounts can be understood as pointing to the fact that we must have something like a background capacities that can be understood in terms of the informational system. If such capacities were not in the Background, there is no reason that we would find it 'obvious' that we are not referring to some causally isolated hermit by "Thales," no matter how many of our Thales-beliefs are true of him. Indeed, it is just such reactions that constitute our having a commitment to the information system. As a result, we could argue that everybody already 'knew' some version of Kripke's account, and they just needed his help, through a *Meno*-like process, to 'recollect' it.

A similar set of Background capacities involving how we understand the language we share with others can be taken to be implicit in our practice. Not only do we have an implicit understanding of how perception and memory are causally related to the world around us, we have an implicit understanding of how language and linguistic communication function, an understanding that involves our speaking shared languages with shared meanings. Our implicit commitment to this model is manifested in not only our content attributions, but also our deference behavior.[448] For example, when our use is corrected, our feeling of having made a certain sort of mistake (misusing a word, saying something false (as opposed to using a non-standard word and saying something

---

[447] I.e., a set of practices that could be explicitly represented as such a system.

[448] If socially infected content ascriptions are justified by the presence of such a background picture of language and such a background picture is manifested in, say, the type of deference behavior described above, then it is of considerable importance that we do defer in such a fashion. Explanations of deference in terms of, say, a pragmatic desire to communicate effectively, if true, would suggest that there is no such background picture of a shared language, in which case it would be extremely difficult to avoid some type of ascriptional individualism. These issues are discussed in fuller detail in 1.222.

misleading)) is explained by, and is partially constitutive of, our having such a picture of language in the background. Kripke's claim that "when the name 'is passed from link to link,' the receiver of the name must … intend when he learns it to use it with the same reference as the man from whom he heard it"[449] can be understood as a somewhat voluntaristic characterization of aspects of this Background understanding of language. With this Background in place, many 'non-individualistic' ascriptions are subject to methodologically individualistic (and in some sense implicitly metalinguistic) explanation without running into the type of objection that Burge lodges against straight metalinguistic accounts. Namely:

> There appears to be a general presumption that a person is reasoning at the object level, other things being equal. The basis for this presumption is that metalinguistic reasoning requires a certain self-consciousness about one's words and social institutions. This sort of sophistication emerged rather late in human history. (Cf. any history of linguistics.) Semantical notions were a part of this sophistication. (Burge, 1979, 97)

The idea is *not* that the intentional content of my "arthritis" belief is spelled out in terms of whatever the community which is causally responsible for my use of "arthritis" means by the term (as it would be on the simple metalinguistic account). Rather the idea is that the intentional content simply involves arthritis, but that it does so in virtue of a non-intentional background that implicitly has a picture in place that has such a result as its consequence.

Searle, of course, does not see things this way and tries to put the Background directly into explicitly metalinguistic intentional contents when he claims that "Often, in fact, one does make what I called parasitic reference using a proper name: often the only identifying description one associates with a name '*N*' is simply the 'object called *N* in my community or by my interlocutors.'"[450] In just the way his account of the content of visual experience did, these formulations neglect the point that the Background should be in the *background*. Such formulations, in addition to being implausibly metalinguistic, are susceptible to various types of counterexamples (e.g., the generally accepted usage has changed since I learnt the name). Much the same can be said of his attempt to deal with the Putnam cases

---

[449] Kripke 1972, 96.
[450] Searle 1983, 243. See also Searle 1983, 244, 250.

with claims such as "The indexical definitions given by Jones on earth of "water" can be analyzed as follows: "water" is defined indexically as whatever is identical in structure with the stuff causing *this* visual experience, whatever that structure is."[451] This, once again, builds too much Background into the content and is also subject to obvious counterexamples.

---

[451] Searle 1983, 207.

Bibliography

Anderson and Owens. 1990. *Propositional Attitudes, The Role of Content in Logic, Language, and Mind*, Stanford: CSLI.

Appiah, A. 1985. *Assertion and Conditionals*, London: CUP.

Aristotle. 1985. *Nicomachean Ethics*. (Erwin trans.), Indianapolis: Hackett.

Austin, J.L. 1956. "A Plea for Excuses", in Austin 1961.

Austin, J.L. 1961. *Philosophical Papers*, Oxford: OUP.

Barrett, R. and Gibson, R. 1990. *Perspectives on Quine*, Cambridge: Blackwell.

Belnap, N. and Green, M. 1994. "Indeterminism and the Thin Red Line", in Tomberlin (ed.) *Philosophical Perspectives, 8: Logic and Language*, Atascadero: Ridgeview.

Bennett, J. 1976. *Linguistic Behaviour*, New York: CUP.

Bilgrami, A. 1992a. *Belief and Meaning*, Cambridge: Blackwell.

Bilgrami, A. 1992b. "Can Externalism be Reconciled with Self-Knowledge?", *Philosophical Topics*, v.20, n.1.

Bilgrami, A. 1993a. "Comments on Chomsky", in Chomsky 1993.

Bilgrami, A. 1993b. "Norms and Meaning", in Stoecker 1993.

Black, M. (ed.). 1962. *The Importance of Language*, Englewood Cliffs: Prentice Hall.

Block, N. 1981. *Readings in the Philosophy of Psychology, volume 2*, Cambridge: MIT.

Block, N. 1986. "Advertisement for a Semantics for Psychology", in French, Uehling & Wettstein. 1986.

Bogdan, R. (ed). 1986. *Belief: Form, content and function*, New York: OUP.

Boghossian, P. 1989. "The Rule Following Considerations", *Mind* Vol. 98, 507-549.

Boyd, R. 1973, "Realism, Underdetermination, and a Causal Theory of Evidence", *Nous* 7: 1-12.

Brandom, R. 1989. *Twelve Lectures on Hegel*, MS, University of Pittsburgh.

Brandom, R. 1994. *Making it Explicit*, Cambridge: Harvard University Press.

Burge, T. 1977. "Belief De Re", *The Journal of Philosophy*. vol. 74.

Burge, T. 1979. "Individualism and the Mental", in French, Uehling & Wettstein 1979b.

Burge, T. 1986a. "Individualism and Psychology", *The Philosophical Review* Vol. 95.

Burge, T. 1986b. "Cartesian Error and the Objectivity of Perception", in Pettit & McDowell, 1986.

Burge, T. 1986c. "Intelectual norms and the Foundations of Mind", *The Journal of Philosophy*. Dec 1986.

Burge, T. 1988. "Individualism and Self-Knowledge",*The Journal of Philosophy*. 1988.

Burge, T. 1989. "Wherein is Language Social?", in George 1989.

Burge, T. 1992. "Philosophy of Language and Mind: 1959-1990", *The Philosophical Review* Vol. 101, no. 1.

Camp, J. 1987. "Why attributions of aboutness report soft facts",*Philosophical Topics* 15.

Chomsky, N. 1986. *Knowledge of Language*, New York. Praeger.

Chomsky, N. 1988. *Language and the Problem of Knowledge*, Cambridge: MIT.

Churchland, P. 1981. "Eliminative Materialism and Propositional Attitudes", *The Journal of Philosophy*. Feb 1981.

Chomsky, N. 1993. *Language and Thought*, London: Moyer Bell.

Collins, A. 1987. *The Nature of Mental Things*, Notre Dame: University of Notre Dame Press.

Davidson, D. 1968. "On Saying That", in Davidson 1984.

Davidson, D. 1974a. "Belief and the Basis of Meaning", in Davidson 1984.

Davidson, D. 1974b. "On the Very Idea of a Conceptual Scheme", in Davidson 1984.

Davidson, D. 1975. "Thought and Talk", in Davidson, 1984.

Davidson, D. 1977. "Reality Without Reference", in Davidson, 1984.

Davidson, D. 1979. "The Inscrutability of Reference", in Davidson, 1984.

Davidson, D. 1982. "Communication and Convention", in Davidson, 1984.

Davidson, D. 1984. *Inquiries into Truth and Interpretation*, Oxford: OUP.

Davidson, D. 1986a. "A Nice Derangement of Epitaphs", in Lepore 1986.

Davidson, D. 1986b. "A Coherence Theory of Truth and Knowledge", in Lepore 1986.

Davidson, D. 1987. "Knowing One's Own Mind", *Proceedings and Addresses of the American Philosophical Association*.

Davidson, D. 1989. "The Myth of the Subjective", in Krausz 1989.

Davidson, D. 1990. "Meaning Truth and Evidence", in Barrett and Gibson 1990.

Davidson, D. 1990b. "The Structure and Content of Truth", *The Journal of Philosophy*, Vol. LXXXVII, no.6.

Davidson, D. 1991a. "What is Present to the Mind?", in Villanueva 1991.

Davidson, D. 1991b. "Epistemology Externalized", *Dialectica*, Vol. 45, N. 2-3.

Davidson, D. 1992. "The Second Person", in French Uehling and Wettstein 1992.

Davidson, D. 1993a. "Reply to Akeel Bilgrami", in Stoecker 1993.

Davidson, D. 1993b. "Reply to Andreas Kemmerling", in Stoecker 1993.

Davidson, D. 1994a. "The Social Aspect of Language", in McGuinness and Oliveri 1994.

Davidson, D. 1994b. "Donald Davidson", in Guttenplan 1994.

Davidson, D. and Harman, G. 1972. *Semantics of Natural Languages*. Boston: D. Reidel.

Davidson, D. and Hintikka, J. 1968. *Words and Objections*. New York: Humanities Press.

Dennett, D. 1982. "Beyond Belief", in Dennett 1987.

Dennett, D. 1987. *The Intentional Stance*. Cambridge: MIT.

Derrida, J. 1977. "Limited Inc a b c . . ." in Derrida 1988.

Derrida, J. 1988. *Limited Inc*. Evanston: Northwestern University Press.

Devitt, M. 1980. *Designation*, New York: Columbia UP.

Devitt, M. 1994. "The Methodology of Naturalistic Semantics", *The Journal of Philosophy*, Oct. 94, 545-572.

Devitt, M. 1996. *Coming to Our Senses*, New York: CUP.

Devitt, M. and Sterelny, K. 1987. *Language and Reality*, Oxford: Blackwell.

Donnellan, K. 1972. "Proper Names and Identifying Descriptions", in Davidson and Harman 1972.

Donnellan, K. 1990. "Belief and the identity of reference", in Anderson & Owens. 1990.

Dreyfus, H. 1991. *Being-in-the-World*, Cambridge: MIT.

Dreyfus, H. and Rabinow, P. 1983. *Michel Foucault: Beyond Structuralism and Hermeneutics*, Chicago: University of Chicago Press.

Dummett, M. 1973. *Frege, Philosophy of Language*, Cambridge: Harvard University Press.

Dummett, M. 1975. "What is a Theory of Meaning? (I)", in Dummett 1993.

Dummett, M. 1991. *The Logical Basis of Metaphysics*, Cambridge: Harvard University Press.

Dummett, M. 1993. *The Seas of Language*, New York: OUP.

Dupré, J. 1981. "Natural Kinds and Biological Taxa", *The Philosophical Review* Vol.XC, No.1, 66-90.

Dworkin, R. 1977. *Taking Rights Seriously*, Cambridge: Harvard University Press.

Eco, U. Santambrogio, M. and Violi, P. 1988. *Meaning and Mental Representations*. Indianapolis: Indiana UP.

Eddington, A. S. 1929. *The Nature of the Physical World*, New York: Macmillan.

Evans, G. 1973. "The Causal Theory of Names", in Evans 1985.

Evans, G. 1982. *The Varieties of Referrence*, New York: OUP.

Evans, G. 1985. *Collected Papers*, New York: OUP.

Feigl, H. Scriven, M. and Maxwell, G. (eds). 1958. *Minnesota Studies in the Philosophy of Science II: Concepts, Theories and the Mind-Body Problem*. Minneapolis, University of Minnesota Press.

Feyerabend, P.K. 1975. *Against Method*. Thetford: Thetford Press Limited.

Field, H. 1972. "Tarski's Theory of Truth", in Platts, 1980.

Field, H. 1973. "Theory Change and the Indeterminacy of Reference", *The Journal of Philosophy* Vol. LXX, August,1973 , 462-481.

Field, H. 1974. "Quine and the Correspondence Theory", *The Philosophical Review*, vol. 83: 200-28.

Field, H. 1978. "Mental Representation", in Block 1981.

Falvey, K. and Owens, J. 1994. "Externalism, Self-Knowledge, and Skepticism", *The Philosophical Review*, vol 103, no. 1.

Flew, A. (ed) 1956. *Essays in Conceptual analysis*, New York: Macmillan.

Fodor, J. 1981. *Representations*. Cambridge: MIT.

Fodor, J. 1987. *Psychosemantics*, Cambridge: MIT.

Fodor, J. 1994. *The Elm and the Expert*, Cambridge: MIT.

Fodor, J. and Lepore, E. 1992. *Holism: a shopper's guide*, Cambridge: Blackwell.

Fodor, J. and LePore, E. 1993. "Why Meaning (Probably) isn't Conceptual Role", in Villanueva 1993.

Frege, G. 1892. "On Sense and Meaning", in Frege 1984.

Frege, G.1891. "Function and Concept", in Frege 1984.

Frege, G. 1984. *Collected Papers*, Oxford: Blackwell.

French, Uehling & Wettstein. 1979a. *Contemporary Perspectives in The Philosophy of Language*, Minneapolis, University of Minnesota Press.

French, Uehling & Wettstein. 1979b. *Midwest Studies in Philosophy IV: Studies in Metaphysics*, Minneapolis, University of Minnesota Press.

French, Uehling & Wettstein. 1986. *Midwest Studies in Philosophy X: Studies in the Philosophy of Mind*, Minneapolis, University of Minnesota Press.

French Uehling and Wettstein. 1992. *Midwest Studies in Philosophy XVII: The Wittgenstein Legacy*, Notre Dame, University of Notre Dame Press.

Gallie, W. B. 1956. "Essentially Contested Concepts", in Black 1962.

Geach, P. T. 1972. *Logic Matters*, Oxford: Blackwell.

Gemes, K. 1991. "The Indeterminacy Thesis Reformulated", *The Journal of Philosophy* vol. LXXXVIII, n. 2.

George, A. (ed). 1989. *Reflections on Chomsky*, New York: Blackwell.

Goodman, N. 1954, 1983. *Fact, Fiction and Forecast*, fourth edition, Cambridge: Harvard UP.

Gould, S. J. 1983. "What, If anything, Is a Zebra?", in his *Hen's Teeth and Horse's Toes*, New York: Norton.

Grandy, Richard. 1973. "Reference Meaning and Belief", *The Journal of Philosophy* Volume 71, 1973, 439-452.

Grice, P. 1957. "Meaning", in Grice 1989.

Grice, P. 1968. "Utterer's Meaning, Sentence-Meaning, and Word-Meaning", in Grice 1989.

Grice, P. 1969. "Utterer's Meaning and Intentions", in Grice 1989.

Grice, P. 1982. "Meaning Revisited", in Grice 1989.

Grice, P. 1987. "Retrospective Epilogue", in Grice 1989.

Grice, P. 1989. *Studies in the Way of Words*, Cambridge: Harvard University Press.

Grimm, R & Merrill, D. 1988. *Contents of Thought*, Tuscon: University of Arizona Press.

Gunderson, K. (ed.). 1975. *Minnesota Studies in the Philosophy of Science VII*, Minneapolis: University of Minnesota Press.

Guttenplan, S. (ed). 1994. *A Companion to the Philosophy of Mind*, Cambridge: Blackwell.

Hegel, G.W.F. 1807, 1977. *The Phenomenology of Spirit,* New York: OUP.

Hobbes, T. 1651, 1981. *Leviathan*, New York: Penguine.

Hylton, P. 1982. "Analyticity and the indeterminacy of translation", *Synthese* 52, 162-84.

Jackendoff, R. 1988. "Conceptual Semantics", In Eco, Santambrogio and Violi 1988.

Jackendoff, R. 1990. *Semantic Structures*, Cambridge: MIT.

Jackman, H. (forthcoming). "Radical Interpretation and the Permutation Principle", forthcoming in *Erkenntnis.*

Kant, I. 1787, 1929. *Critique of Pure Reason*, London: Macmillan.

Kaplan, D. 1968. "Quantifying in", in Davidson and Hintikka, 1968.

Katz, J. 1975. "Logic and Language: An Examination of Recent Criticisms of Intentionalism", in Gunderson 1975.

Kay, P. 1983. "Linguistic Competence and Folk Theories of Language", *Proceedings of the Berkeley Linguistics Society*, 1983.

Kirk, R. 1986. *Translation Determined*, New York: OUP.

Krausz, M. (ed) 1989. *Relativism: Interpretation and Confrontation*, Notre Damme: Notre DammeUniversity Press.

Kripke, S. 1972. *Naming and Necessity*, Cambridge: Harvard University Press.

Kripke, S. 1977. "Speaker's Reference and Semantic Reference", in French et all, 1979a.

Kripke, S. 1979. "A Puzzle about Belief", in Salmon and Soames 1988.

Kripke, S. 1980. "Preface" to Kripke 1972.

Kripke, S. 1982. *Wittgenstein on Rules and Private Language*, Cambridge: Harvard University Press.

Kuhn, T. 1973. "Objectivity, Value Judgment, and Theory Choice", in Kuhn 1977.

Kuhn, T. 1977. *The Essential Tension*, Chicago: The University of Chicago Press.

Lakoff, G. 1987. *Women, Fire and Dangerous Things*, Chicago: The University of Chicago Press.

Lepore, E, (ed). 1986. *Truth and Interpretation,* Oxford: Blackwell.

Lepore, E, and Van Gulick, R, (eds). 1991. *John Searle and His Critics,* Oxford: Blackwell.

Lewis, C.I. 1929. *Mind and the World Order*, New York: Dover.

Lewis, D. 1969.*Convention*, Cambridge: Harvard University Press.

Lewis, D. 1974. "Radical Interpretation", in Lewis 1983b.

Lewis, D. 1983a. "New Work for a Theory of Universals", *Australasian Journal of Philosophy*, v.61, n.4.

Lewis, D. 1983b. *Philosophical Papers, Volume 1*. New York: OUP.

Loar, B. 1985. "Social Content and Psychological Content", in Grimm & Merrill 1988.

Locke, J. 1975. *An Essay Concerning Human Understanding*, New York: OUP.

Lyas, C. (ed) 1971. *Philosophy and Linguistics*, New York: Macmillan.

Lycan, W. 1984. *Logical Form in Natural Language*, Cambridge: MIT.

Marconi, D. 1995. "On the Structure of Lexical Competence", *PAS* vol. XCV part 2.

Mates, B. 1958. "On the Verification of Statements about Ordinary Language", in Lyas 1971.

McDowell, J. 1986. "Singular Thought and the Extent of Inner Space", in Pettit & McDowell, 1986.

McGinn, C. 1982. "The Structure of Content", in Woodfield 1982a.

McGuinness and Oliveri (eds.). 1994. *The Philosophy of Michael Dummett*, Dordrecht: Kluwer.

Mill, J. S. 1843. *A System of Logic*, New York: Harper Brothers.

Melville, H. 1983. *Moby Dick*, New York: The Library of America.

Nagel, T. 1986.*The View from Nowhere*, New York: OUP.

Paprzycka, K. forthcoming. PhD thesis, University of Pittsburgh.

Peirce, C. S. 1878. "How to Make our Ideas Clear", in Peirce 1986.

Peirce, C. S. 1986. *Writings of Charles S. Peirce, Volume 3: 1872 - 1878*, Bloomington: Indiana University Press.

Pettit, P, and McDowell, J, (eds). 1986a. *Subject, Thought, and Context*, New York: OUP.

Pettit, P, and McDowell, J. 1986b. "Introduction", to Pettit and McDowell 1986a.

Platts, M. (ed). 1980. *Reference, Truth and Reality*, London: Routledge & Kegan Paul.

Prior, A. 1967. *Past, Present and Future*, Oxford: OUP.

Putnam, H. 1975a. "The Meaning of 'Meaning'", in Putnam 1975b.

Putnam, H. 1975b. *Mind Language and Reality*, New York: CUP.

Putnam, H. 1981a. "Reference and Understanding", in Putnam 1981b.

Putnam, H. 1981b. *Meaning and the Moral Sciences*, Boston: Routledge & Kegan Paul.

Putnam, H. 1981c. *Reason, Truth and History*, New York: CUP.

Putnam, H. 1981d. "Why there isn't a Ready-Made World", in Putnam 1985.

Putnam, H. 1985 . *Realism and Reason*, New York: CUP.

Putnam, H. 1988. *Representation and Reality*, Cambridge: MIT.

Putnam, H. 1993. "Aristotle after Wittgenstein", In Putnam 1994.

Putnam, H. 1994. *Words and Life*, Cambridge: Harvard University Press.

Quine, W.V. 1951. "To Dogmas of Empiricism", in Quine 1953.

Quine, W.V. 1953. *From a Logical Point of View*, Cambridge: Harvard University Press.

Quine, W.V.O. 1960. *Word and Object*, Cambridge: MIT.

Quine, W.V. 1975. "On Empirically Equivalent Systems of the World", *Erkenntnis* 9 : 313-328.

Quine, W.V.O. 1979. "Facts of the Matter", in Shahan and Swoyer 1979.

Quine, W.V.O. 1987. "Indeterminacy of Translation Again", *The Journal of Philosophy*, vol. LXXXIV, no. 1.

Quine, W.V.O. 1990. *The Pursuit of Truth* , Cambridge: Harvard University Press.

Rawls, J. 1971. *A Theory of Justice*, Cambridge: Harvard University Press.

Rosch, E. 1975. "Family resemblances: studies in the internal structure of catagories", *Cognitive Psychology*, 7 .

Rosenberg, J. 1974 *Linguistic Representation*, Dordrecht: D. Reidel.

Rouse, J. 1987. *Knowledge and Power*, Ithaca: Cornell University Press.

Russell, B. 1905. "On Denoting", in Russell 1956.

Russell, B. 1918. "The Philosophy of Logical Atomism", in Russell 1956.

Russell, B. 1956. *Logic and Knowledge*, London: Unwin.

Salmon, N. 1982. *Reference and Essence*, Oxford: Blackwell.

Salmon, N. 1986. *Frege's Puzzle*, Cambridge: MIT.

Salmon, N and Soames, S. (eds). 1988. *Propositions and Attitudes*, New York: OUP.

Schiffer, S. 1972. *Meaning*, Oxford: OUP.

Searle, J. 1958. "Proper Names", in Strawson 1967.

Searle, J. 1970. *Speech Acts*, New York: CUP.

Searle, J. 1977. "Reiterating the differences: a reply to Derrida", *Glyph* 2.

Searle, J. 1979. *Expression and Meaning*, New York: CUP.

Searle, J. 1983. *Intentionality*, New York: CUP.

Searle, J. 1987. "Indeterminacy, Empiricism, and the First Person", *The Journal of Philosophy* Vol. LXXXIV, no.3, 123-146.

Searle, J. 1991. "Responce: Reference and Intentionality", in Lepore and Van Gulick 1991.

Searle, J. 1992. *The Rediscovery of the Mind*, Cambridge: MIT.

Sellars, W. 1954. "Some Reflections on Language Games", in Sellars 1963.

Sellars, W. 1963. *Science Perception and Reality,* London, RKP.

Sellars, W, and Chisholm, R. 1958. "Correspondence on Intentionality", in Feigl, Scriven and Maxwell, 1958

Shahan and Swoyer (eds.). 1979. *Essays in the Philosophy of V.W. Quine*, Hassocks: Harvester.

Sperber & Wilson. 1986. *Relevance*, Cambridge, Harvard UP.

Stalnaker, R. 1984. *Inquiry*, Cambridge: MIT.

Stebbing, L. S. 1937. *Philosophy and the Physicists*, London: Methuen.

Stich, S. 1983. *From Folk Psychology to Cognitive Science*, Cambridge: MIT.

Stich, S. 1984. "Relativism, Rationality and the Limits of Intentional Description", *Pacific Philosophical Quarterly* 65, 211-235.

Stich, S. 1986. "Are Belief Predicates Systematically Ambigous?", in Bogdan 1986.

Stich, S. 1990. *The Fragmentation of Reason*, Cambridge: MIT.

Stoecker, R. (ed). 1993. *Reflecting Davidson*, New York: de Gruyter.

Strawson, P.F. 1964. "Intention and Convention in Speech Acts", in Strawson 1971.

Strawson, P. F. (ed). 1967. *Philosophical Logic*, New York: OUP.

Strawson, P.F. 1971. *Logico-Linguistic Papers*, London: Methuen.

Thomason, R. 1984. "Combinations of Tense and Modality", in *Handbook of Philosophical Logic*, vol II, Dordrecht: D. Reidel.

Travis, C. 1995. "Order out of Messes: Critical Notice of Bilgrami's *Belief and Meaning*", *Mind* v. 104.

Urmson, J. O. 1953. "Some Questions Concerning Validity", In Flew 1956.

Vellanueva, E. 1991. *Philosophical Issues 1: Consciousness*, Atascadero: Ridgeview.

Vellanueva, E. 1993. *Philosophical Issues 3: Science and Knowledge*, Atascadero: Ridgeview.

Wilson, M. 1982. "Predicate Meets Property", *The Philosophical Review*, XCI, No. 4.

Wittgenstein, L. 1953. *Philosophical Investigations*, third edition. Oxford: Blackwell.

Woodfield, A. (ed). 1982a. *Thought and Object*, New York: OUP.

Woodfield, A. 1982b. "Introduction", in Woodfield 1982a.