

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

¹⁰⁹ Kripke 1979, 113.

¹¹⁰ Kripke, 1979, 113.

¹¹¹ This thought of Dummett's is put in an especially vivid form in Evans 1982, 67-9.¹¹² And the Gricean analysis is by no means the only one from which it follows (see section 3.1).

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 wastepaper 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 Registering 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 accounts¹²¹

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 through 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 Gricean 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' accounts¹²⁷ '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 acquires 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' solution¹³⁴ 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 one 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 robust¹³⁵ 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. Sellars¹⁴⁴ and Sellarsians such as Brandom¹⁴⁵ 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 Burge¹⁴⁶ and the interpretation of Wittgenstein advocated by Kripke¹⁴⁷ 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 advocacy 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 an intermediate position will not lose 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 to 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 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.