

Intention and Convention in the Theory of Meaning

Stephen Schiffer

New York University

What is the relation between language and thought—or, more exactly, between the representational, or intentional, characteristics of language and those of propositional attitudes such as believing and intending? An answer many find attractive is that we must “explicate the intentional characteristics of language by reference to believing and to other psychological attitudes.”¹ In other words, it is only our intentional mental states—believing, intending, and the like—that have *original* intentionality, intentionality that doesn’t have its source in something else’s intentionality; the intentionality of words and speech acts is *derived* intentionality, intentionality inherited from that of associated mental states. That answer raises two questions: how is the original intentionality of thought to be explained, and how does the intentionality of language “derive” from the original intentionality of thought? The focus of this chapter will be on the second question, although the first question can’t be ignored entirely, and I shall touch on it in the final section of this chapter.

Although many philosophers seem to accept the derivation view, hardly any of them attempt to spell out how it works. In fact, I can think of only two programs that attempt to explain how the intentionality of language reduces to that of thought, and only one of these programs ventures to reduce all questions about linguistic representation to questions about mental representation. The more ambitious program is one that derives from the Grice-inspired program of intention-based semantics (IBS); the other derives from David Lewis’s project of defining what I shall call the *public-language relation*. I shall start with a reconstruction of Lewis’s account of the relation in *Convention*² because a problem that immediately arises for that account provides a natural segue to the more ambitious IBS project.

¹ Chisholm (1958), cited in Speaks (2009).

² Lewis (1969).

I. *Lewis on the Public-Language Relation*

Hardly any philosopher of language would deny that if something is an expression which has meaning in a population, then that is by virtue of facts about the linguistic behavior and psychological states of members of that population. Philosophers of language would like to know which facts those are. They would also like to know what it is for something to be the language of a population. The two questions are apt to seem very closely related, for it's apt to seem that an expression has a certain meaning in a population just in case it has that meaning in the language of that population. If one both thinks of a language as a "pairing of sound and meaning over an infinite domain"³ and accepts that an expression e has meaning just in case there is something x such that e means x , then one may think of a language as a function—doubtless a finitely specifiable function—that maps finite sequences of sounds or marks (or whatever) onto meanings (or, to accommodate ambiguity, sets of meanings). Then, if L is such a function, we may say, first, that

$$e \text{ means* } x \text{ in } L \text{ iff } L(e) = x$$

(where '*' is to remind us that this stipulated sense of 'means*' isn't the use-dependent notion of meaning that philosophers struggle to understand), and then, second, that

$$e \text{ means } x \text{ in } P \text{ iff, for some } L, L \text{ is a language of } P \text{ and } e \text{ means* } x \text{ in}$$

L .

A language, thus conceived, pairs the words, phrases and sentences of a language onto the things they mean* in the language. That is bound to seem inevitable, for isn't the meaning of a sentence determined by the meanings of its constituent morphemes in conjunction with the semantic import of the syntactic structures deployed in the sentence? And doesn't that thought lead inexorably to the thought that an account of expression-meaning must first say what it is for a morpheme to have a certain meaning and what it is for a syntactic structure to have a certain semantic import, and to do that in a way that will determine a meaning for every expression of the language? If that is one's sense of how things must be, then the project of saying what it is for a function of the kind just described to be the language of a population ought to strike one as bewilderingly complex. Where is one even to begin, and how could one possibly complete the

³ Chomsky (2006: 92).

project without being able, first, to specify a generative grammar for the language, where that is a finitely specifiable theory of the language that generates one or more syntactic structures for each expression of the language and interprets those structures both phonologically and semantically, and then, second, to specify the myriad interlocking practices whereby the morphemes and structures of the language would come to mean in the population what by stipulation they mean* in the language? Anyone who has banged her head against the apparent inevitability, and then the apparent impossibility, of that approach must appreciate the genius and simplicity of David Lewis's way of cutting through that Gordian knot.

In presenting the views I will discuss in this chapter my focus will be on their essential plot lines, and I will simplify like crazy in order not to get bogged down at every turn with technical complexities. To that end I will pretend that the languages we speak have no indexicality, ambiguity or moods other than the indicative, and that the meaning of a sentence is a proposition, in the generic sense of an abstract, mind- and language-independent entity that has a truth condition, which it has both necessarily and absolutely (i.e. without relativization to anything else).⁴ Relative to those simplifications, I will say that a *Lewis-language* is any function L from finite sequences of sounds or marks (or whatever)—the “sentences” of L —onto propositions. This allows us to say not only that a sequence σ means* the proposition q in the Lewis-language L just in case $L(\sigma) = q$, but also that:

For any Lewis-language L and sequence of sounds σ , σ is *true* in L iff
for some q , σ means* q in L and q is true.

Now a language may be used in any number of ways. For example, a language whose “sentences” are sequences of neural activity may function as a person's language of thought. David Lewis's interest in *Convention* is in a population's using a language as a public language of communication, a language they use to communicate with one another, and his book aims to say what relation must hold between a Lewis-language L and a population P in order for it to be the case that P uses L as a public language of communication. Let's call that relation, whatever it turns out to be, the *public-language relation*. If L is the language members of P use as their medium of communication, then every sentence of L will mean in P what it means* in L . Since the meaning a sentence has in a population is the use-dependent notion of sentence-meaning

⁴ See Schiffer (2003: Ch. 1in).

philosophers want to understand, an account of the public-language relation would be an account of the use on which a sentence's meaning depends, provided that the languages we speak really are Lewis-languages. If they are, and if the only intentionality involved in the account of the public-language relation is that of propositional-attitudes, then the account would have succeeded in defining the intentionality of sentences in terms of the intentionality of thought.

Lewis's definition of the public-language relation in *Convention* is the following, minus an addendum which I will get to presently:

For any Lewis-language L and population P , L is a public language of P iff there prevails in P a convention of truthfulness in L .

Roughly speaking, a convention for Lewis is a regularity in behavior to which the members of a population want to conform if (nearly) everyone else in the population conforms, and to which they do conform because it's common knowledge among them that they expect one another to conform. Such regularities are self-perpetuating in that past conformity gives rise to the expectation of conformity, which gives rise to future conformity.

In *Convention* Lewis said that

It is *common knowledge* in a population P that ___ if and only some state of affairs A holds such that:

- (1) Everyone in P has reason to believe that A holds.
- (2) A indicates to everyone in P that everyone in P has reason to believe that A holds.
- (3) A indicates to everyone in P that ____.⁵

I independently introduced a similar notion in *Meaning*,⁶ which I called *mutual knowledge*,⁷ and said that x and y mutually know q just in case x knows q , y knows q , x knows that y knows q , y knows that x knows q , x knows that y knows that x knows q , and so on, and I proposed finite

⁵ Lewis (1969: 56).

⁶ Schiffer (1972).

⁷ Actually, I called the notion mutual knowledge* to make clear that I was stipulatively defining a technical notion, rather than trying to define a notion that was already called 'mutual knowledge' in the vernacular. It's not surprising Lewis and I independently hit on essentially the same notion, since we each invoked the notion to preclude essentially the same kind of counterexample, although for Lewis the counterexample he needed to prevent was to the account of convention he was developing, whereas the counterexample I needed to prevent was to the Gricean account of speaker-meaning I was trying to repair.

conditions for the generation of mutual knowledge. The generalization to the n -person case is obvious, but for mutual knowledge in a population whose members aren't all acquainted with one another, I in effect proposed that q is mutual knowledge in P just in case everyone in P knows q , everyone in P knows that everyone in P knows q , and so on.⁸ In his (1975) Lewis revised his account of common knowledge and said that a proposition is "*common (or mutual) knowledge* [in P just in case it] is known to everyone [in P], it is known to everyone [in P] that it is known to everyone [in P], and so on," but he added that the knowledge may be "merely potential: knowledge that would be available if one bothered to think hard enough." Conformity to a convention of truthfulness in L requires one not to utter any sentence of L unless it's true in L , but because "truthfulness-by-silence is truthfulness,"⁹ it's not enough for a convention of truthfulness in L to prevail in P that members of P never utter sentences of L ; in order for a convention of truthfulness to prevail in P members of P must regularly utter sentences of L and (for the most part) utter them only when they believe them to be true.

A striking feature of Lewis's account of the public-language relation is that, while it yields a definition of a sentence's having a certain meaning in a population, it says nothing at all about the meanings of words, nor does it say anything about a sentence's meaning being determined by the meanings of its constituent words and the semantic import of its syntactical construction. Lewis of course is well aware of the fact that:

Not just any arbitrary infinite set of verbal expressions will do as the set of sentences of an interesting language. No language adequate to the purposes of its users can be finite; but any language usable by finite human beings must be the next best thing: finitely specifiable. It must have a finite grammar, so that all its sentences, with their interpretations, can be specified by reference to finitely many elementary constituents and finitely many operations for building larger constituents from smaller ones.¹⁰

He also of course recognizes that words as well as sentences have meanings for those who use them, and that the meaning of a word is determined, insofar as it is determined, by the way those

⁸ Schiffer (1972: § II.2).

⁹ Lewis (1969: 165-6).

¹⁰ Lewis (1969: 166-6).

for whom it has meaning use it. Why then doesn't Lewis define language in a way that requires a language to have a grammar, so that if a language L is used by a population P , then L 's words will have in P whatever meanings the grammar for L assigns them? Lewis explains that, while a grammar for L uniquely determines L , L doesn't uniquely determine any grammar: more than one grammar will determine L , grammars that may differ in the meanings they assign to the morphemes of L or even in what they recognize to be the morphemes of L . This presents Lewis with a problem:

Given P , we select L by looking for a convention of truthfulness; but given L , how can we select [its grammar]? Conventions of truthfulness pertain to whole sentences and leave the interpretations of parts of sentences undetermined. Perhaps we should look for conventions of some other kind, but I cannot think what the content of such a convention might be.¹¹

Well, perhaps the correct grammar for L is determined by something other than a convention. Lewis considers the Chomskian conjecture that the correct grammar for a language is the one that enters into the explanation of the linguistic competence of its users, but he rejects using that psycholinguistic hypothesis in defining the public-language relation; for even if it's true, he says, it's a contingent truth, and thus can't be part of an *analysis* of ' L is used by P ', "since the analysandum clearly could be true although the analysans was false."¹² Lewis is forced to conclude that he knows "of no promising way to make objective sense of the assertion that a grammar Γ is used by a population P whereas another grammar Γ' , which generates the same language as Γ , is not."¹³

As I have so far presented Lewis, one may be puzzled as to why Lewis thinks grammars pose a problem for him. Why not identify languages with what he calls grammars, namely, finite specifications of the functions he now calls languages? Such a grammar may be conceived, at least initially, as a function that maps each expression of the language it determines onto its meaning in the language, and does so in a way that reveals how the meaning of every semantically complex expression is determined by its syntactical construction and the meanings

¹¹ Op. cit.: 198.

¹² Lewis (1975: 178).

¹³ Ibid.: 177.

the function assigns to the morphemes from which the complex expressions are constructed. Let L_G be such a Lewis-language-cum-grammar. In the functions Lewis defines as languages, each sequence of sounds in the function's domain is a sentence of the language. The sequences of sounds that constitute the domain of L_G will include words and other sub-sentential expressions along with sentences. If, for any ϵ and μ , $L_G(\epsilon) = \mu$, then we may say that ϵ means* μ in L_G , and if L_G is a public language of population P , then ϵ means μ in P . For Lewis, a grammar determines the language of which it's a grammar, but a language doesn't determine its grammar. Since L_G is what Lewis calls a grammar, we would have a conception of language that does determine its grammar. Now suppose Lewis were to say that L_G is a public language of P just in case there prevails in P a convention of truthfulness in L_G , where, as before, there prevails in P a convention of truthfulness in L_G just in case it's common knowledge in P that its members try not to utter any *sentence* of L_G unless it's true in L_G , etc. Then, provided the languages we speak are Lewis-languages, we can say what it is for a word w to mean μ in P : it means that just in case for some language L_G , L_G is a public language of P , w is a word in L_G and $L_G(w) = \mu$. A Lewis-language maps only sentences onto meanings; let's call a finitely specifiable function L_G which maps expressions of every syntactic category onto a meaning a *Chomsky-language*. Every Chomsky-language determines a unique Lewis-language, but a Lewis-language determines no unique Chomsky-language. If there prevails in P a convention of truthfulness in a Lewis language, then we have no way of saying that any grammar of the language is the grammar used in P . But if there prevails in P a convention of truthfulness in a Chomsky language, then it follows that a particular grammar for a particular Lewis language is used in P . We get that result simply by virtue of the fact that all it takes, according to Lewis, for there to be a convention of truthfulness in L_G in P is that it be common knowledge in P that, for any sentence σ of L_G and proposition q , members of P try not to utter σ unless $L_G(\sigma) = q$ and q is true, etc.

Lewis didn't overlook this easy way of securing that a Chomsky language is used in a population: he had already ruled it out by a stipulation he had made along the way about how the just-cited common knowledge condition was to be understood. Lewis says that a function L from finite sequences of sounds or marks—the “sentences” of L —into propositions is a public language of a population P only if it is common knowledge in P that members of P try never to utter a sentence of L unless it's true in L . But wouldn't such knowledge require ordinary speakers to know propositions of the form $L(\sigma) = q$, and wouldn't that require those ordinary

speakers to know some set theory? And, just as bad, wouldn't knowing such propositions require them to have a finitely specifiable way of thinking about the function L , and wouldn't that be tantamount to knowing a grammar for L ? Lewis has no doubt that ordinary speakers have no such knowledge, and to avoid having it required by the common knowledge required for them to use a language, he first distinguishes between two ways in which the common knowledge in question might obtain in:

It's common knowledge in P *in sensu composito* that there prevails in P a convention of truthfulness in L iff it's common knowledge in P that for any σ, q such that $L(\sigma) = q$, a member of P won't utter σ unless she thinks q is true.

It's common knowledge in P *in sensu diviso* that there prevails in P a convention of truthfulness in L iff for any σ, q such that $L(\sigma) = q$, it's common knowledge in P that a member of P won't utter σ unless she thinks q is true.¹⁴

In other words, when the common knowledge is *in sensu composito*, what is known requires members of P to have a conception of L and enough knowledge of set theory to know propositions of the form $L(\sigma) = q$, whereas when the common knowledge is *in sensu diviso* members of P needn't have any way of thinking of L or any knowledge of set theory; they merely have to have the right expectations when sentences of L are uttered, so that if $L(\sigma) = q$ and a member of P hears another member of P utter σ , then she will expect him to believe q . "The common man," Lewis says, "need not have any concept of L in order to expect his fellows to be truthful ... in L . He need only have suitable particular expectations about how they might act ... in various situations."¹⁵

Now that we know why Lewis thinks he can say what it is for a sentence, but not for a word, to mean something in a population, and now that we have a better understanding of what,

¹⁴ The generality *in sensu composito/diviso* distinction is from Abelard (Lewis 1969: 64); those familiar with the knowledge *de dicto/de re* distinction will recognize that Abelard's distinction is definable in terms of that distinction.

¹⁵ Lewis (1975: 180).

according to him, must be the case in order for a convention of truthfulness to prevail in a population, we are positioned to assess his claim that:

For any Lewis-language L and population P , L is a public language of P iff there prevails in P a convention of truthfulness in L .

There are, I believe, at least two problems with this account.¹⁶ The first is that if there prevails in a population a convention of truthfulness in any language, then there will also prevail in that population conventions of truthfulness in infinitely many languages that are not public languages of the population. Suppose, for example, that English is the public language of P . Then members of P will regularly utter sentences of English, and when they do they will, for the most part, believe those sentences to be true. But only a finite number of the infinitely many English sentences will ever be uttered. Suppose ξ is an English sentence so convoluted and long that no one could reasonably expect it ever to be uttered. Then it will be obvious to any member of P who considers ξ that no member of P would utter it, and, *a fortiori*, for any proposition r , obvious that no member of P will utter ξ unless r is true. As noted earlier, with respect to the sentences of a population's language that will never be uttered, the truthfulness that obtains is truthfulness-by-silence. But now let Gobbledygook be a language that coincides with English with respect to every sentence that might be uttered but departs wildly from English thereafter: the sentences in Gobbledygook but not in English may be composed of words that aren't in English, or, if Gobbledygook and English have the same sentences, then the sentences that no one would ever utter have meanings in Gobbledygook that are entirely different from the meanings they have in English. Since it will be common knowledge (*in sensu diviso*) in P that members of P regularly utter sentences of Gobbledygook and that when they do they intend to be truthful in Gobbledygook and that they will be truthful-by-silence as regards the sentences of Gobbledygook that they know no one in P would ever utter, it follows by Lewis's definitions that there prevails in P a convention of truthfulness in Gobbledygook—a language that is not a

¹⁶ There is a third problem I feel compelled to mention although I can't hope to develop it here (I do try to develop it in "How Vagueness Affects Meaning," an unpublished MS that is still a work in progress). The problem is that a language can't be described as a pairing of sounds and meanings unless there are such things as meanings—that is to say, unless an expression's having meaning consists in there being some thing that it means. It is my view that that isn't what having meaning consists in.

public language of P .¹⁷ And since infinitely many languages satisfy the description of Gobbledygook, it follows from Lewis's definitions that infinitely many languages constitute counterexamples to his definition of the public-language relation.

I mentioned this problem to Lewis in 1968; he acknowledged that it showed that his definition failed to provide a sufficient condition, but he couldn't at the time find a revision that avoided the problem. That may seem surprising. The counterexample works only if one takes '¬ A unless B ' to be equivalent to the material condition ' $A \rightarrow B$ ', which one can know to be true merely by knowing that A is false. But if Lewis were to say instead that the common knowledge is of the *counterfactual* proposition *that if a member of P were to utter ξ , then she would think that q was true*, then Gobbledygook wouldn't be a counterexample to the definition, for if, for any σ , if $G(\sigma) \neq E(\sigma)$, then no member of P would expect a member of P to mean $G(\sigma)$ if she were to utter σ , for they wouldn't know to associate that proposition with σ . The problem with this thought, however, is that there are also infinitely many English sentences that members of P wouldn't be able understand—sentences like 'Buffalo buffalo buffalo buffalo buffalo buffalo buffalo',¹⁸ in addition to the infinitely many English sentences whose length or convoluted structure make them impossible to process. In "Language and Languages,"¹⁹ Lewis proposed a certain way of avoiding the problem, but then in "Meaning without Use: Reply to Hawthorne,"²⁰ he recognized that that solution to what he now called the *meaning-without-use problem* didn't work, and proposed a third account of the public-language relation that, as we'll see in §IV, may also be problematic.

The second objection to *Convention's* account of the public-language relation also shows that that account fails to provide a sufficient condition for a language to be a public language of a population. The problem is that if it were a sufficient condition for L 's being a public language of P that there prevailed in P a convention of truthfulness in L , then virtually *every* convention

¹⁷ Lewis (Ibid.: 187) gives a pithy restatement of the problem: "A sentence never uttered at all is *a fortiori* never uttered untruthfully. So truthfulness-as-usual in [English] plus truthfulness-by-silence on the garbage sentences constitutes a kind of truthfulness in [Gobbledygook] ... Therefore we have a prevailing regularity of truthfulness ... in [Gobbledygook]. This regularity qualifies as a convention in P ..."

¹⁸ Pinker (1992). Pinker attributes the example to Annie Senghas.

¹⁹ Lewis (1975).

²⁰ Lewis (1992).

would count as a convention of truthfulness in a language, and this by virtue of the fact that every (or virtually every) convention requires certain actions to be performed when certain conditions obtain. For example, suppose that in monastery M there is a convention to recite prayer A on Monday, prayer B on Wednesday, and prayer C on Friday, and not to recite those prayers on any other days. Now let L^\dagger be that function such that:

$\forall x,y[L^\dagger(x) = y$ iff (1) $x =$ reciting A & $y =$ the proposition that it's Monday or (2) $x =$ reciting B & $y =$ the proposition that it's Wednesday or (3) $x =$ reciting C & $y =$ the proposition that it's Friday]

Then L^\dagger is a Lewis-language and there prevails in M a convention of truthfulness in L^\dagger , and Lewis is committed to saying that L^\dagger is used in M as a public language of communication. But since the members of M never use L^\dagger to communicate anything (at least to one another), it's clearly not used as a public language of communication (or as any other kind of language). I mentioned this problem, too, to Lewis in 1968; he agreed that it was a counterexample to his definition and changed that definition to:

For any Lewis-language L and population P , L is a public language of P iff there prevails in P a convention of truthfulness in L , *sustained by an interest in communication*,

which is the definition that appeared in *Convention* when the book was published. The revision, however, doesn't do the job: prayers A , B and C might all be for greater powers of communication; yet that still wouldn't make L^\dagger a language the monastery uses as a public language of communication. A convention must forge a considerably tighter connection to communication if it's to succeed in defining the public-language relation; it would at least have to secure that sentences of the language it concerns are uttered in order for speakers to mean what the sentences mean* in the language. This is our segue to the Gricean program of intention-based semantics.

II. *Intention-Based Semantics*

This is a program for reducing all questions about the intentionality of speech acts and linguistic expressions to questions about the intentionality of thought. It takes as foundational in the theory of meaning a certain notion of *speaker-meaning* and seeks to define it, without recourse to any

semantic notions, in terms of acting with certain audience-directed intentions. Then it seeks to define other agent-semantic notions—most notably, speaker-reference (the notion of a *speaker*'s referring to a thing, as contrasted with an *expression*'s referring to it) and illocutionary acts²¹—in terms of its defined notion of speaker-meaning.²² With that done, Gricean IBS then sets out to define the semantic features of linguistic expressions wholly in terms of its defined notion of speaker-meaning, together with ancillary notions, such as that of *convention*, which are themselves explicable wholly in terms of non-semantic propositional attitudes. Since expression-meaning is defined in terms of speaker-meaning and convention, and speaker-meaning and convention are defined in terms of non-semantic propositional attitudes, it's supposed to follow that expression-meaning is also defined in terms of non-semantic propositional attitudes.

There is disagreement among Griceans as to how exactly the definition of speaker-meaning should go, but there may be little reason to care about how to define a notion of speaker-meaning if that notion can't be used to define the semantic features of linguistic expressions, and in this regard a reader of Grice's 1957 article "Meaning" has a right to be puzzled. Grice spends nearly all of that article building up to his famous proposal that *S* meant something in "uttering" *x* iff *S* "intended the utterance of *x* to produce some effect in an audience by means of the recognition of this intention,"²³ and it is only at the very end of the article that we get anything about expression-meaning. But then all we get are two equivalences baldly presented without any elaboration. The first is that:

‘*x* meant something’ is (roughly) equivalent to ‘Somebody meant something [in uttering] *x*’.

For example, if Gretel rolls her eyes to communicate to Hansel that the speaker is a pretentious bore, then, by the definition, Gretel's eye rolling meant something. The second generalization is that:

²¹ The notion of an illocutionary act—certain acts we perform in uttering sentences, such as telling, asking, requesting, ordering, warning, and so on—is due to Austin (1962).

²² For attempts to define illocutionary acts in terms of a Gricean notion of speaker-meaning, see Strawson (1964) and Schiffer (1972: IV). For an attempt to define speaker-reference in terms of Gricean speaker-meaning, see Schiffer (1978) and (1981).

²³ Grice (1957: 385). Grice uses ‘utterance’ and its cognates in a technical sense that includes non-linguistic items and behavior.

‘ x means (timeless) that so-and-so’ might as a first shot be equated with some statement or disjunction of statements about what “people” (vague) intend (with qualifications about “recognition”) to effect by x .

There are four things one is apt to find puzzling. First, Grice has made no effort to show how these definitions are motivated by his definition of speaker-meaning. Second, the equivalence offered for ‘ x meant something’ seems not to be an analysis of any obvious pre-theoretic notion, but is more in the nature of a stipulation whose theoretical purpose hasn’t been revealed. Third, there is no mention of word meaning. And fourth, the second equivalence, which is intended to cover indicative sentence meaning, appears to ignore the fact that every natural language has infinitely many sentences that will never be, or even could be, uttered. So why is Grice’s article so famous and thought by many to be of such importance? Does the importance of Grice’s article reside wholly in his suggested account of speaker-meaning, never mind any relevance that account might have for an account of expression-meaning?

No; those who perceived Grice’s article to be important did so because they took themselves to discern in his account of speaker-meaning an *invisible hand* that guides them from the intentions that define speaker-meaning to an account of expression-meaning in terms of those intentions. Any Gricean account of speaker-meaning could be used to make the invisible hand visible, but I will use the account of assertoric speaker-meaning implicit in Grice’s 1957 article. According to that account:

For any person S , proposition p and utterance x , S meant p in uttering x iff, for some person A , S uttered x intending

- (1) A to believe p ;
- (2) A to recognize that S uttered x intending (1);
- (3) A ’s recognition of that intention to function as part of A ’s reason for believing p .

Quite apart from the question of whether this is in any way “correct,” its proper understanding requires seeing the answers Grice assumed to two questions raised by the definition. One question was how A ’s recognition of S ’s intention to get A to believe p was supposed to function as part of A ’s reason for believing p . The intended answer was that A would infer from the fact that S uttered x intending A to believe p that S believed she knew p , and then, taking the fact that

S believed she knew p to be very good evidence that p was true, infer p from that fact.²⁴ The other question raised by the above displayed definition was how A was to recognize that S uttered x intending A to believe p . Understanding how that recognition was supposed to work is essential to understanding the invisible hand that was supposed to guide one from the account of speaker-meaning to an account of expression-meaning. It was supposed to work like this: when S utters x in order to mean p , there is some feature φ such that S intends x to have φ and intends A to recognize that x has φ and to infer in part therefrom that S uttered intending A to believe p . Making this explicit yields the following slightly tweaked version of Grice’s account of assertoric speaker-meaning:

- For any person S , proposition p and utterance x , S meant p in uttering x iff for some feature φ and person A , S uttered x intending
- (1) x to have φ ;
 - (2) A to recognize that x has φ ;
 - (3) A ’s recognition that x has φ to function as at least part of A ’s reason for believing that S uttered x intending:
 - (4) A to believe p ;
 - (5) A ’s recognition of S ’s intending A to believe p to function as at least part of A ’s reason believing p .

Let’s call the value of ‘ φ ’ in an act of speaker-meaning its *inference-base feature* (its *IB-feature*, for short). Acts of speaker-meaning are typically performed by uttering sentences of a language common to the speaker and her audience, and the IB-features of those sentences are their meanings. If you utter ‘It’s snowing’ to communicate that it’s snowing, the IB-feature of the sentence ‘It’s snowing’ on which you rely is the meaning of that sentence in English. The

²⁴ There is an obvious problem for the Gricean account of assertoric speaker-meaning given that this is how recognition of intention is supposed to result in A ’s believing the proposition S uttered x intending A to believe—namely, that while (3) might be a necessary condition for S ’s *telling* A p , it’s not a necessary condition for S ’s *meaning* p ; for if it were a necessary condition, then Grice would not have meant anything in writing any of the sentences in “Meaning,” and this because, while Grice produced those sentences intending us to believe what they expressed, he certainly did not intend his readers to believe what he wrote on his authority; that is to say, he didn’t intend his readers to believe anything he said because Grice’s believing it was good evidence of its truth.

reason IB-features are typically meaning properties is that they are optimal IB-features: if you want to tell your child that it's snowing, you would do much better to utter 'It's snowing' than to attempt to communicate that it's snowing by impersonating a snow flake or uttering 'The flamingoes are flying south early this year'. At the same time, a *sine qua non* of a Gricean account of speaker-meaning is that the only intentional notions mentioned on its right-hand side are ordinary propositional-attitude notions, and, consequently, it's not a necessary condition for a person's meaning a proposition that what she utters have a semantic property as its IB-feature. For example, during a lecture one might communicate to one's friend that one is bored by closing one's eyes and pretending to snore. The invisible-hand in the displayed account of speaker-meaning is that in any population whose members have frequent need to communicate with one another, the fact that they frequently communicate with one another will result in their utterances having IB-features that are both optimal and specifiable wholly in terms of what they know their speaker-meaning practices to be. The invisible-hand idea is that, since the notion of speaker-meaning utilized in these optimal IB-features is defined in wholly non-semantic terms, these features will also be intrinsically specifiable in wholly non-semantic terms. The Gricean's invisible-hand strategy is completed by *identifying* meaning properties with those non-semantically specifiable optimal IB-features, thereby explaining the optimality of meaning properties as IB-features.

The Gricean account of assertoric speaker-meaning requires one further tweak before we have an account of assertoric speaker-meaning that best reveals the Gricean strategy for explaining expression-meaning in terms of the conditions that define speaker-meaning. Earlier (on p. 000) I mentioned the notion of *common knowledge* Lewis introduced in *Convention* and the similar notion of *mutual knowledge* that I introduced in *Meaning*. The further tweak is that mutual (or common) knowledge must be added to the Gricean mix to yield this account of speaker-meaning:

[SM] For any person S , proposition p and utterance x , S meant p in uttering x iff for some feature φ and person A , S uttered x intending it to be mutual knowledge between S and A that x has φ and, at least partly on that basis, mutual knowledge that S uttered x intending A to believe p and intending their

mutual knowledge that S uttered x intending A to believe p
to be at least part of A 's reason for believing p .

Mutual knowledge was originally introduced to repair the failure of Grice's original conditions to provide a set of jointly sufficient conditions for speaker-meaning;²⁵ I invoke it now because of the way it will be needed in the Gricean attempt to define expression-meaning in terms of speaker-meaning. There are better and worse ways to understand mutual knowledge, and I might not have opted for the best way in my book. The essential job mutual knowledge needs to perform is to capture the sense in which acts of communication require the defining features of speaker-meaning to be "out in the open" between speaker and hearer. In *Meaning* I offered a set of finite conditions for generating mutual knowledge, and I now think that I would have done best simply to have identified mutual knowledge with a version of those base conditions. In any case, for present purposes I'll continue to use 'mutual knowledge' and its cognates as dummy expressions for whatever turns out to be the best accommodation of the requisite out-in-the-openness.

The Gricean takes his invisible-hand strategy to be most clearly and paradigmatically exhibited in his account of simple signals,²⁶ and a little thought experiment will show how that is supposed to work. There is a weekly seminar regularly attended by the same people. Their practice is to raise their hands if they want to be called on, but, while they would benefit from one, they have no simple way of indicating that what they want to contribute is a follow-up question. During one session, a visitor from the University of Latvia, Zuzka, raises her hand during a lively discussion and moves her index finger rapidly up and down. It's clear to all that Zuzka intends to communicate something by this gesture, but at first no one can figure out what it is. After several minutes it somehow transpires that in Latvian universities that sort of finger movement means that one has a follow-up question. Now suppose that during another exchange the following week one of the attendees, Harvey, raises his hand and moves his index finger rapidly up and down. In this case, everyone in the class will know straightway that Harvey means that he has a follow-up question. What explains this dramatic difference? Why was it

²⁵ See e.g. *ibid.*: II. 1 & 2.

²⁶ Relative to ongoing simplifying assumptions, γ is *simple signal* in population P just in case, for some q , γ means q in P , but there are no constituents of γ such that γ 's meaning q is a function of the meanings of those constituents.

that no one knew what Zuzka meant in moving her finger in way Ω , whereas one week later the very same people effortlessly and immediately knew that in moving his finger in way Ω Harvey meant that he had a follow up question? The answer, of course, is that at the time Zuzka performed the finger movement it had no feature that was an effective IB-feature in the seminar for meaning that one had a follow-up question, but after that the movement had a few features it didn't previously have, and these features separately and together constituted quite an effective IB-feature for meaning that one had a follow-up question. One of these features was that of being mutually known to be such that it was performed by Zuzka in her attempt to communicate to the seminar that she had a follow-up question; another was that of being mutually known to be the standard way in Latvian universities to communicate *that one had a follow-up question* (Q , for short). Since it benefited the seminar to have a simple way to indicate that one had a follow-up question, now that no one doubts what one would mean by moving one's index finger in way Ω , it's apt to catch on, so that it soon becomes mutual knowledge in the seminar that there is a practice of meaning Q by moving one's index finger in way Ω , and that mutual knowledge makes the gesture an optimal IB-feature as regards meaning Q in the seminar.

You may recognize that we have entered the territory of the kind of self-perpetuating regularities that David Lewis showed to be conventions, and at this point it seems correct to say that the gesture means Q in the seminar. Now the gesture Ω doesn't mean a proposition, since each person who makes the gesture will mean that *she* has a follow-up question. But to keep what is essential to the lines we are exploring from being hidden in complexities and qualifications we are ignoring non-assertoric speech acts, indexicality and ambiguity, and pretending that sentence-size meanings are propositions. Relative to all that, the Zuzka example suggests an account of simple-signal meaning that entails the following:

[SIMP] For any x , proposition q and population P , x is a simple signal that means q in P iff it's mutual knowledge in P that there is a practice in P of meaning q by uttering x and intending that mutual knowledge to function as x 's IB-feature when a member of P means q by uttering x .

Practices that satisfy SIMP are conventional practices,²⁷ but it may not be easy to describe them as conventions. At least SIMP doesn't require there to be a convention in P to mean q by uttering x , which is good, since x can mean q in P even though members of P have ways other than uttering x to mean q , nor does it require there to be a convention in P not to utter x unless one thereby means q , which is also a good thing, for even if x means q in P members of P might communicate other propositions by using x metaphorically. In any case, whether or not the practice entailed by SIMP can be described as a convention, as opposed to a conventional practice, that wouldn't give us reason to suppose SIMP fails to provide either a necessary or a sufficient condition for simple-signal meaning.

The Gricean sees an even broader application for the kind of IB-feature employed in SIMP. Recall that for the Gricean speaker-reference is to be defined in terms of speaker-meaning,²⁸ and this will be so in a way that suggests that the name-of relation is best captured by something along the lines of:

[N] n is a name of y in P iff it's mutual knowledge in P that there is a practice in P of speakers' referring to y with n & intending that mutual knowledge to be n 's IB-feature.²⁹

In other words, when you say to me 'Saul Kripke is giving a talk today' you intend the feature of 'Saul Kripke' that enables me to know that you are referring to Saul Kripke to be that it's mutual knowledge in a population to which we both belong that there is a practice of referring to Kripke with 'Saul Kripke', which for the Gricean is roughly equivalent to saying that what enables me to know that you are referring to Kripke with 'Saul Kripke' is that it's mutual knowledge between us that 'Saul Kripke' is the name of Saul Kripke in a population to which we both belong.

Of course, the Gricean can't hope to account for the meanings of natural language sentences in a similar way, since a sentence has its meaning even if no one has ever uttered it. How, then, might the Gricean invisible-hand strategy apply to natural languages? The Gricean expects that if σ means q in the language L of a population P then, while there needn't be any

²⁷ See Schiffer (1972: V.3).

²⁸ See Schiffer (1978), (1981) and (2015).

²⁹ Cf. the account of the name-of relation in Evans (1973), which is based on the account of simple signals in Schiffer (1972).

practice in P of uttering σ , there will prevail in P a set of practices pertaining to L such that one utters σ in conformity with those practices only if one means q in uttering σ . The question is, what might those practices be? Unfortunately for the Gricean, we are already positioned to see that his goal of defining expression-meaning in terms of his defined notion of speaker-meaning is an impossible goal for him to achieve. The invisible-hand strategy fails to take one to any good place. We can appreciate this in the following way.

SIMP, the lately displayed Gricean definition of simple-signal meaning, is apt to seem plausible. The only semantic notion used in SIMP is the notion of speaker-meaning, and SM, which I'm taking to represent the Gricean account of speaker-meaning, has been defined in terms of non-semantic propositional attitudes. It may therefore seem that, if both SIMP and SM are correct, then the Gricean has succeeded in defining simple-signal meaning in wholly non-semantic terms, and thereby at least to have shown that the intentionality of simple signals derives from the intentionality of the propositional attitudes in terms of which it's defined. *Not so!* Even if SIMP and SM are correct, that would not entail that simple-signal meaning had been indirectly defined in terms that included those that define speaker-meaning. The point is familiar and simple. Suppose that

$$A =_{\text{def}} \dots X \dots Y \dots$$

and that

$$B =_{\text{def}} \dots A \dots Z \dots$$

then

$$B =_{\text{def}} \dots X \dots Y \dots Z \dots$$

does *not* follow if the context of 'A' in ' $B =_{\text{def}} \dots A \dots Z \dots$ ' is intentional. Specifications of knowledge are intentional contexts *par excellence*, and *speaker-meaning is mentioned in SIMP only in specifying the mutual knowledge required for simple-signal meaning*. Consequently, we have no reason to expect that if SIMP is correct, then SIMP_G, the result of replacing every speaker-meaning expression in SIMP with what SM provides as its definitional expansion, will also be correct. That it would be wrong to expect SIMP_G to be true if SIMP is true would seem to be confirmed by a side-by-side comparison of the two definitions:

| SIMP | SIMP _G |
|--|--|
| <p>For any x, proposition q and population P, x is a simple signal that means q in P iff it's mutual knowledge in P that there is a practice in P of meaning q by uttering x and intending that mutual knowledge to function as x's IB-feature when a member of P means q by uttering x.</p> | <p>For any x, proposition q and population P, x is a simple signal that means q in P iff it's mutual knowledge in P that there is a practice in P whereby a member S of P utters x in order that there be something y that S utters such that, for some feature φ and member A of P, S utters y intending it to be mutual knowledge between A and S that y has φ and, at least partly on that basis, mutual knowledge that S uttered y intending A to believe q and further intending their mutual knowledge that S uttered y intending A to believe q to be at least part of A's reason for believing q, and in such utterances of x S intends the inference-base feature φ of x to be the fact that it's mutual knowledge in P that there is a practice in P whereby a member S of P utters x in order that there be something y that S utters such that, for some feature φ and member A of P, S utters y intending it to be mutual knowledge between A and S that y has φ and, at least partly on that basis, mutual knowledge that S uttered y intending A to believe q and further intending their mutual knowledge that S uttered y intending A to believe q.</p> |

SIMP is apt to seem plausible, SIMP_G quite implausible, which isn't surprising, given that we have no reason at all to expect the two definitions to have the same truth-value even if SIMP and SM are both true.³⁰

The Gricean invisible-hand strategy for defining the semantic properties of expressions in non-semantic terms never gets off the ground: the Gricean simply has no way of defining expression-meaning in terms of the *intentions* that by his lights are constitutive of speaker-meaning. But that doesn't preclude defining expression-meaning in terms of the *concept* of speaker-meaning, and that suggests the possibility of a convention-based way of reducing the semantic to the psychological that isn't strictly Gricean *intention*-based semantics but might be considered a cousin of it and, if correct, would show how the intentionality of language derives from the intentionality of thought. The next section explores this idea.

III. *An Almost-Gricean Semantics*

For all we yet know, SIMP, which defines simple-signal meaning in terms of a certain kind of conventional speaker-meaning practice, remains plausible. All that was shown was that the attempt to replace the speaker-meaning expressions in SIMP with their Gricean expansions results in an implausible account of simple-signal meaning, even if the Gricean account of speaker-meaning is correct. One familiar diagnosis of this substitutivity failure which derives from Frege is that when an expression occurs in an intentional context, as when it's used to specify what someone knows, the expression doesn't denote the object, property or relation it denotes when it occurs in extensional contexts, but denotes instead a *concept* of what it denotes in those extensional contexts. The idea, in other words, is that the occurrence of 'means' in the sentence 'In making that gesture Zuzka meant that she had a follow-up question' denotes the speaker-meaning relation, but its occurrence in an intentional context, such as the context created by 'knows' in 'Sid knows that in making that gesture Zuzka meant that she had a follow-up question', denotes a *concept* of that relation and not the relation itself. Suppose that, or something enough like it, is right. Then we would have defined simple-signal meaning not, as the Gricean would have it, in terms of the intentions that define speaker-meaning, but rather in

³⁰ An even better demonstration of the point being made is obtained by replacing the speaker-reference expressions in *N*, the plausible-looking definition of the name-of relation, with the definition of speaker-reference in terms of speaker-meaning I offered in (1981).

terms of what, for all we yet knew, was a primitive concept of speaker-meaning.³¹ But *if* (i) that concept of speaker-meaning could be identified with a psychological construct that was itself definable in non-intentional terms—say, in terms of the inferential and causal roles of neural expressions in mentalese, (ii) the speaker-meaning relation it denoted was definable without recourse to any intentionality other than the intentionality of non-semantic propositional attitudes, and (iii) we can move beyond SIMP to give an account of the semantic properties of all expressions in terms of conventional practices whose specification involved that concept of speaker-meaning, *then* we could achieve something like a reduction of the semantic to the psychological. And if, further, the speaker-meaning relation denoted by the concept of speaker-meaning was definable *à la* Grice and meaning properties could still be conceived as IB-properties, then—what the hell—we would have a theory that was close enough to Gricean semantics to be called an almost-Gricean semantics.

To see whether that can be done, a good place to begin would be to assume that SIMP, the account of simple-signal meaning, was correct and try to see whether speaker-meaning conventions (or conventional practices) might fare better than conventions of truthfulness in a definition of the public-language relation. We are encouraged to think that might be so by the problem with which we left Lewis’s proposal that

For any Lewis-language L and population P , L is a public language of P iff there prevails in P a convention of truthfulness in L , sustained by an interest in communication.

This was the problem that virtually every convention—e.g. a convention to drive on the right or to wear casual clothes to work on Fridays—was for Lewis a convention of truthfulness in a “language” that was in no sense used as a medium of communication, not even when the convention was somehow or other “sustained by an interest in communication.” The encouragement this problem gave to a Gricean approach was that it suggested that in order for L to count as a public language of P the convention (or conventional practice) governing the use of L should be a convention conformity to which requires uttering sentences of L in order to mean what those sentences mean* in L . In pursuing this line of thought, however, it’s important to keep in mind that, while a correct definition of the public-language relation would *eo ipso* be an

³¹ See Schiffer (1987: § 9.3).

account of what it was for a sentence to have meaning in a population, that would be no stopping point even for almost-Griceans unless the account of the public-language relation, unlike those Lewis had on offer, somehow or other managed also to define what it was for an expression of *any* syntactic kind—words and phrases, as well as sentences—to have meaning in a population.

The most obvious Gricean counterpart to Lewis’s notion of a convention of truthfulness would be a convention of meaning, where that is taken to mean that there prevails in P a convention not to utter any sentence σ of L unless, for some q , $L(\sigma) = q$ and in uttering σ one means q , and where that in turn entails mutual knowledge which, when understood *in sensu diviso*,³² comes to:

- (A) For any Lewis-language L and population P , L is a public language of P iff for any σ , q , if $L(\sigma) = q$, then it’s mutual knowledge in P
- (1) that a member of P won’t produce an unembedded utterance of σ unless she means q thereby;³³

³² We saw below (p. 000) that it’s essential that the knowledge of L be *in sensu diviso*, lest speakers of a language be required to have something they clearly don’t have—to wit, knowledge of the syntactic and semantic rules of L . Grice’s own suggestion about how to define expression-meaning falls afoul of this requirement. His suggestion, put in the terms of this chapter, is that a sentence σ means q in P just in case for any (or nearly any) member S of P , S has a “resultant procedure” to utter σ if S wants to mean q , where a resultant procedure for σ would be a procedure that results from the procedures S has for uttering the constituents and structures of σ . Now that doesn’t yet require S to have propositional knowledge of what his procedures are; the problem arises when we ask how another member of P , A , is to know what S means when she utters σ , for here Grice says that this is to be accomplished via A ’s knowledge that S has in her repertoire the procedure of uttering σ if she wants to mean q . But S can hardly intend A to know that σ has the IB-feature of *being such that S has in her repertoire the resultant procedure of uttering σ if she wants mean q* unless she intends A to compute that σ has that feature via her knowledge of the procedures S has in her repertoire for the morphemes and structures from which σ is composed, and it’s impossible to see how that is to be accomplished without knowledge of the language’s semantics and syntax. It will also be notice that Grice’s claim *that a sentence σ means q for S only if S has in her repertoire the resultant procedure of uttering σ if she wants to mean q* also runs afoul of the meaning-without-use problem.

³³ An occurrence of a sentence is unembedded just in case the occurrence is not within some other sentence.

- (2) that if a member of P means q in uttering σ , then she intends σ 's IB-feature to be the fact that it's mutual knowledge in P that (1).

(A) has more than a few problems, the most worrisome of which are, first, that it suffers from the same meaning-without-use problem that (A)'s counterpart in Lewis was seen to suffer from and, second, that it doesn't explain what it is for a word or other subsentential expression to have meaning in a population. Brian Loar did as much as any Gricean to understand expression-meaning in Gricean terms, and he offered an account of the public-language relation that both took into account the problems that arose for Lewis's account and provided an account of what it is for an expression of any syntactic kind to have meaning in a population.³⁴ Taking our cue from Loar, and ignoring some fixable problems, we need to take three more steps to achieve what by Loar's lights would be a correct account of the public-language relation (relative to our ongoing simplifications about ambiguity, vagueness, indexicality, non-indicative moods and every sentence having a proposition as its meaning).

The first step is to counterfactualize the mutual knowledge in (A), thereby getting:

- (B) For any Lewis-language L and population P , L is a public language of P iff for any σ , q , if $L(\sigma) = q$, then it's mutual knowledge in P
- (1) that if a member of P were to produce an unembedded utterance of σ she would mean q thereby;
- (2) that if a member of P means q in uttering σ , then she intends σ 's IB-feature to be the fact that it's mutual knowledge in P that (1).

The problem with (B), as Loar noticed, is one already encountered in our discussion of Lewis: it requires speakers of a language to know the meaning of every sentence of their language, whereas we already know that there are infinitely many sentences of every spoken language that are too long or convoluted for ordinary speakers to understand. The fix Loar proposes, which takes us to the second step, restricts the mutual knowledge to those sentences of their language members of a population can understand:

³⁴ Loar (1976).

- (C) For any Lewis-language L and population P , L is a public language of P iff *there is a large enough restriction L' of L such that, for any σ, q , if $L'(\sigma) = q$, then it's mutual knowledge in P*
- (1) that if a member of P were to produce an unembedded utterance of σ she would mean q thereby;
 - (2) that if a member of P means q in uttering σ , then she intends σ 's IB-feature to be the fact that it's mutual knowledge in P that (1).

Loar knew that (C) was no stopping point, for it has the same meaning-without-use problem that Lewis's definition of the public-language relation has: infinitely many languages that aren't public languages of any population will each entail the restriction L' before differing wildly from one another beyond what they have in common; and, of course, it doesn't explain what constitutes a word's having meaning in a population.

The languages we are trying to capture are finitely specifiable. A finite specification of a language is a grammar. In *Convention* Lewis argued that, if the infinite language L is the language of P , then L must be finitely specifiable, but there will be no fact of the matter as to which finite specification of L is the correct specification of L . Lewis reverses himself in (1992). There he comes to the realization that in order to determine which language is a population's public language it must be possible to determine which grammar generates that language, but he now thinks he knows how this can be done:

True, there are many grammars. But they are not on equal terms. Some are "straight" grammars; for example, any grammar that any linguist would actually propose. Others are "bent," or "gruesome," grammars; for example, what you get by starting with a straight grammar for English and adding one extra rule, which states that every expression with more than forty occurrences of the word 'cabbage' is a sentence meaning that God is great. We have no difficulty in telling

the difference We can reasonably hope that all straight grammars that agree on the used fragment will agree everywhere.³⁵

The application of this solution to (C) yields:

(D) For any Lewis-language L and population P , L is a public language of P iff there is a large enough restriction L' of L and a grammar Γ such that (a) Γ is the “straightest” grammar that determines L' , (b) Γ determines L , and (c) for any σ , q , if $L'(\sigma) = q$, then it’s mutual knowledge in P

(1) that if a member of P were to produce an unembedded utterance of σ she would mean q thereby;

(2) that if a member of P means q in uttering σ , then she intends σ ’s IB-feature to be the fact that it’s mutual knowledge in P that (1).

A striking feature of this solution to the meaning-without-use problem is that it doesn’t require the grammar that determines a person’s language to play any role in the psycholinguistic explanation of the information processing that underlies, and thus accounts for, the person’s ability to understand utterances in that language. The striking fact invites an objection to the solution, which I raised when Lewis and I discussed a version of his (1992) in 1990—namely, that if we learned that the internally represented grammar implicated in the explanation of a person’s ability to understand novel sentences was in fact a bent grammar that determined a language $L^\#$, then we should want to say that $L^\#$ was her language, even if the only straight grammar that fit the used fragment determined a different language. Lewis was unmoved:

Maybe there is a grammar somehow written into the brain. And conceivably it is a bent grammar, so that the language it generates differs, somewhere outside the used fragment, from the language we get by straight extrapolation. Schiffer has asked: does straight extrapolation give the right answers even then? I think so. If not, then whenever we resort to extrapolation to answer questions of syntax and

³⁵ Lewis (1992: 110).

semantics, we are engaged in risky speculation about the secret workings of the brain. That seems wrong.³⁶

Risky speculation about the secret workings of the brain is certainly to be avoided, but, as we'll presently see, it's not entailed by the hypothesis that the criterion for being a population's language is that it's determined by the grammar that explains how members of the population are able to understand the utterances they hear. For suppose that the infinite language L is in fact the public language of population P . Let L' be the fragment of L that has been or might be produced in P . We may assume that each member of P has a compositional understanding of L' , one that relies on the only straight grammar to fit L' and the only grammar that both fits L' and determines L . We should certainly insist that L is the language of P . Now it's possible for there to be another population P^* such that L' is also the fragment of their language that has been or might be produced in P^* , but with this difference: each sentence of L' is for them a *noncomposite* utterance type. In other words, each sentence of L' is for them as a simple signal, such as a fire alarm, is for us: it has propositional meaning, but its meaning is not in any way determined by semantic features of its parts and structure. The "language" L' is simply a large finite set of simple signals; the members of P^* have prodigious memories, and they have learned the sentences of L' the only way they could learn them—namely, one by one. They know what their sentences mean because they have learned what each one means as a single fact, and they have no way of computing what a sentence means on the basis of their knowledge of its syntax and the meanings of its parts, as, from their perspective, a sentence no more has a syntax and semantically relevant parts than a fire alarm has for us. Consequently, members of P^* have no way of understanding a sentence that belongs to L but not to L' ; they have no way, in fact, of determining the meaning of any novel sentence. Clearly, the infinite language L is *not* used in P^* . Yet every straight grammar that generates L' , the language that *is* used by them, is a grammar of L . There are also counterexamples to the straight-grammar solution involving infinite languages. Suppose, for example, that the members of a secret society make up a language whose syntax and semantics they actually write down and formally adopt under a Mafia-like oath, and that, for some reason, perverse or not, the grammar they adopt is by their explicit

³⁶ Lewis (1992: 110, fn. 6).

design a bent grammar. Let's also suppose that an internal representation of the bent grammar plays an essential role in their processing of utterances in their invented language. I would think that the language of the society is the one their bent grammar describes, whether or not there is a straight grammar that fits the sentences the members of the society actually produce or are likely to produce.

Brian Loar has suggested that appealing to a grammar that is somehow written into the brain is exactly what must be done both to solve the meaning-without-use problem and to achieve an account not just of a sentence's meaning in a population but of every expression's meaning in a population. He first stipulates that "*L* is grounded in *P*, with regard to its restriction *L'*, just in case those correlations of sentential features and meaning contributions which figure in the correct psychological explanation of the continuing mastery of *L'* (i.e. effective *L*) by members of *P* will generate, when extended, the full language *L*, including its incomprehensibly complex sentences."³⁷ This would yield:

- (E) For any Lewis-language *L* and population *P*, *L* is a public language of *P* iff there is a large enough restriction *L'* of *L* such that, for any σ, q , if $L'(\sigma) = q$, then (a) it's mutual knowledge in *P*
- (1) that if a member of *P* were to produce an unembedded utterance of σ she would mean *q* thereby;
 - (2) that if a member of *P* means *q* in uttering σ , then she intends σ 's IB-feature to be the fact that it's mutual knowledge in *P* that (1)
- and (b) *L* is grounded in *P* with regard to *L'*.

But what about Lewis's objection that if the grammar that determines the used language is the one written into the brain, then "whenever we resort to extrapolation to answer questions of syntax and semantics, we are engaged in risky speculation about the secret workings of the brain"? Well, suppose I believe that the correct grammar for my language is whatever grammar is written into the brains of those of us who use the language. It would be no small feat to construct a grammar that merely fits the fragment of my language that has been or is ever likely

³⁷ Op. cit.: 159-60.

to be used, but suppose that after considerable labor experts come up with a grammar Γ that gets the fragment right. Γ will determine not just the fragment in question but an entire infinite language: it assigns a syntactic, semantic and phonological interpretation to each of infinitely many sentences, all but a small minority of which have no chance of ever being used by me or anyone else. It seems a very good bet that Γ is a straight grammar, probably the only straight grammar anyone can come up with that fits the used fragment. So now I must decide: Is Γ the grammar implicated in my language processing, or is that grammar one that coincides with Γ on the used fragment but differs from it in that it interprets every sentence containing forty or more occurrences of 'cabbage' as meaning that God is great? I will of course go with the straight grammar Γ , but not on the basis of a risky speculation. My going with the straight grammar Γ is a risky speculation about the secret working of the brain only if the scientist who infers a straight theory from the evidence that underdetermines it, rather than one of the infinitely many bent theories that fits the same evidence, isn't making a risky speculation about the infinitely many possible unexamined cases her theory must cover.

Nevertheless, there is a serious problem with (E) or any other almost-Gricean account of the public-language relation which avoids the meaning-without-use problem by appeal to the grammar implicated in the explanation of our language processing. The problem is that if, like Loar, one invokes groundedness in one's account of the public-language relation—that is to say, into one's account of what determines an expression to have meaning in a population—then one is invoking something that in effect makes the rest of one's account superfluous. For if one is invoking groundedness, then one needs nothing more than:

- (F) For any Lewis-language L and population P , L is a public language of P iff (1) members of P regularly communicate with one another by uttering sentences of L and meaning thereby what those sentences mean* in L , and (2) for some grammar Γ of L , their ability to know what members of P mean in uttering those sentences is grounded in Γ .³⁸

³⁸ This seems very much along the lines of what Stephen Laurence (1996) has proposed as “a Chomskian alternative to convention-based semantics.”

(F) is neither Gricean nor convention-based; it uses the notion of speaker-meaning, but relies on no *particular* account of it, Gricean or otherwise. I'm not suggesting that (F) is correct. My point is merely that it would appear that in order for an almost-Gricean convention-based theorist to have an account of the public-language relation (= an account of expression-meaning) she will in effect need Loar's notion of groundedness, which entails the right-hand side of (F), and that by anyone's lights should already be a necessary and sufficient condition for a Lewis-language's being a public language of a population.³⁹

IV. *What Endures?*

In this chapter I have looked critically at two programs that should dominate any curriculum bearing the label "Intention and Convention in the Theory of Meaning": David Lewis's convention-based account of sentence meaning and the more comprehensive Grice-inspired program of intention-based semantics. As for Lewis, one may object to his account of convention⁴⁰ and one may deny that the languages we speak can be represented as Lewis-languages,⁴¹ but we have seen that even if Lewis's account of convention is correct and the languages we speak can be represented as Lewis-languages, no convention or conventional practice—whether of truthfulness or of speaker-meaning—can explicate what it is for an expression to have meaning for a person or population of persons.

The more ambitious IBS also fails in its attempt to explain expression-meaning. First, it simply doesn't follow that if one can define expression-meaning in terms of speaker-meaning and speaker-meaning in terms of intentions that one can thereby define expression-meaning in terms of intentions. That would follow only if mention of speaker-meaning in the analysis of expression-meaning occurred in a non-intentional context, whereas its occurrence is in a specification of the mutual knowledge the account requires, a paradigm intentional context. Second, as we saw in our discussion of almost-Gricean IBS, an expression's meaning in a population can't be defined in terms of the speaker-meaning practices that prevail in the population. What of the possibility of a Gricean account of speaker-meaning? True, the limitations we now see on the work it could do considerably diminish the interest in having such

³⁹ Much of this paragraph and the one before it is taken from Schiffer (2006).

⁴⁰ See e.g. Gilbert (1981).

⁴¹ See above n. 15.

an account, but nothing has been said so far in this chapter to show that no such account can be correct. And while such an account couldn't be used to explicate expression-meaning, it's not unreasonable to suppose it would have relevance to speech-act theory and to issues about the semantics/pragmatics interface, especially since Grice is nowadays best known for his theory of implicature, and for Grice (nearly enough)

S implicates p in uttering σ iff in uttering σ *S* means but doesn't say *p*,

where in Grice's quasi-technical sense of 'say',

S said p in uttering σ iff *p* "fits" the meaning of σ and *S* meant *p* in uttering σ ,

where "fitting" is exemplified in, say, the way the proposition *that Betty is ready to take the exam* fits the meaning of 'She is ready'. And there remains the unsettled question of whether semantic intentionality can be reduced to the intentionality of mental states.

There are reasons to doubt that a Gricean account of speaker-meaning can be achieved. First, it's not even clear what an account of "speaker-meaning" is supposed to be an account of when it's not constrained by the need to be the speech-act notion needed to define expression-meaning.⁴² Grice's "Meaning" (and, ahem, my own *Meaning*, as well as other early work in Gricean semantics, such as Jonathan Bennett's *Linguistic Behaviour*) gives the impression that the target notion of speaker-meaning is the concept expressed by the ordinary language use of sentences of the form 'In uttering *x* *S* meant that such and such'. However, an attentive reading of Grice (or any other Gricean) reveals that that can't exactly be what is at issue. For example, Gricean accounts of speaker-meaning would approve the form of such speaker-meaning reports as:

- (1) In uttering 'Is Frankfort the capital of Kentucky?', *S* meant *that you were to tell her if Frankfort was the capital of Kentucky*.
- (2) In uttering 'Move to Frankfort, Kentucky!', *S* meant *that you were to move to Frankfort, Kentucky*.

Yet such reports are not likely to be found in ordinary speech. A speaker who uttered 'Is Frankfort the capital of Kentucky?' or 'Move to Frankfort, Kentucky!' would be said to have meant something, but if asked what the speaker meant no ordinary speaker would answer with

⁴² This is elaborated in Schiffer (1982).

(1) or (2). Rather than respond with (1) and (2) she would be much more likely to respond, respectively, with ‘He asked if Frankfort was the capital of Kentucky’ and ‘He told you to move Frankfort, Kentucky’. This doesn’t mean that the Gricean is wrong to use sentences like (1) and (2) as canonical forms for reporting what speakers *mean* when they perform interrogative or imperatival speech acts, but it does seem to show that in using those representations he is stipulating a technical use of the verb ‘to mean’. But then it’s not clear what technical notion is being introduced. The Gricean’s benchmark meaning reports are those he uses to report assertoric acts of speaker-meaning, such as ‘In uttering ‘Il pleut’ Pierre meant that it was raining in Paris’. In such reports the ‘that’-clause is taken to refer to what the speaker meant, which is taken to be a proposition of some stripe or other. This suggests that the Gricean takes assertoric speaker-meaning to be a relation between speakers and propositions of some yet unspecified kind. At the same time, while the ‘that’-clauses in (1) and (2) are being used to specify what the speaker meant, they can’t be referring to propositions, nor, I should think, to anything else. One is left to wonder what univocal notion the Gricean takes speaker-meaning to be. Self-described Griceans today, such as Stephen Neale,⁴³ are unconcerned that no Gricean analysis of speaker-meaning is recognized even by Griceans as being correct. Their commitment seems to be to the idea that what *S* means in uttering a sentence σ is determined entirely by intentions she has in uttering σ , where those intentions are intrinsically specifiable in non-semantic terms, and where the sentence uttered serves only as the means by which *S* intends those intentions to be recognized. I doubt that that idea can be correct. When Jack utters ‘What’s the capital of Kentucky?’ his *intention* is to *ask* Jill what the capital of Kentucky is, and when Jill responds, ‘Frankfort is the capital of Kentucky’, her intention is to *tell* Jack that Frankfort is the capital of Kentucky. I doubt that if we subtract those semantic intentions from all the intentions Jack and Jill had in producing their utterances that there would be enough left over to entail that they performed the speech acts they in fact performed.⁴⁴

Other concerns emanate from *vagueness*. I think one salubrious effect of Gricean work on meaning is a better appreciation of the role of a speaker’s intentions in the determination of what she meant and of what references she made, especially as such intentions may be crucial to

⁴³ See e.g. Neale (1992).

⁴⁴ Schiffer (1987: Ch. 9).

the semantics of indexicals and other referring expressions. At the same time, I think certain features of vagueness show that the extent to which intentions determine what a speaker meant or referred to is quite limited. A single example should make clear what I have in mind. Jane is speaking on the phone with a friend in Los Angeles, and at one point she remarks, ‘It’s raining here’. Let’s suppose that Jane made her remark while standing under an umbrella in pouring rain. Then her utterance is determinately true, and she clearly meant *something* in making her utterance. But *what* did she mean and what *determines* whatever it is that she meant? Well, what she meant is partly determined by the reference of her utterance of ‘here’, so to what location did her utterance of ‘here’ refer? How one answers that question will depend on one’s views about vagueness. The epistemic theorist of vagueness will say that there is some absolutely precise region of space α to which Jane’s utterance of ‘here’ referred. If this is correct, then it should be obvious that that reference wasn’t determined by Jane’s referential intentions, since, having no way to distinguish α from the uncountably many precise regions with which it overlapped, she could not have intended to refer to α . The supervaluationist will say that there is no location that is determinately the referent of her utterance, but there are myriad—indeed, uncountably many—locations each of which is such that it’s indeterminate whether it’s the referent, or indeterminate whether it’s indeterminate whether it’s the referent, and so on (he will say that Jane’s utterance is true because it’s true in each of those “precisifications”). Suppose α is one of the uncountably many precise locations each of which is such that it’s indeterminate whether it’s the referent of Jane’s utterance of ‘here’. Then a theorist who wants to say that a speaker’s referential intentions determine the reference of the indexicals she utters will want to say that it’s indeterminate whether Jane intended to refer to α with her utterance of ‘here’. But it should be obvious that she determinately did *not* intend to refer to α with her utterance of ‘here’.⁴⁵ How could she, when she couldn’t pick out α from the billions of minutely differing overlapping areas if her life depended on it? A theorist who thinks that vagueness isn’t confined to words and concepts might want to say that a certain vague area is the referent of Jane’s utterance of ‘here’. But if α is the vague area to which Jane’s utterance of ‘here’ referred, Jane will be wholly ignorant of that fact: just ask her to tell you which locations are borderline—or borderline borderline, etc.—cases of being included in α . Of course, what

⁴⁵ Cf. Buchanan and Ostertag (2005) and Buchanan (2010).

goes for ‘here’ in Jane’s utterance goes also, *mutatis mutandis*, for what she meant in uttering ‘It’s raining here’. The rub for intention-based semantics is that what goes for Jane’s utterance goes also for virtually *every* utterance; for virtually every sentence uttered is to some extent vague.

What of the view that we must “explicate the intentional characteristics of language by reference to believing and to other psychological attitudes,”⁴⁶ which seemed to motivate much of the work in IBS and arguably in Lewis’s convention-based semantics.⁴⁷ It’s difficult to say in the absence of a clear and plausible account of how exactly the intentionality of language derives from the intentionality of thought, and, most important, it’s difficult to see why we should care whether the view is correct when what is apt to seem to be most important is that the intentionality of thought and language should supervene on non-intentional facts. At one time Brian Loar and I thought it important to *identify* intentional facts with facts that are intrinsically specifiable in non-intentional terms, and we thought that to get such a reduction it was important first to reduce to semantics to psychology, so that one could then reduce psychology to physical and functional facts in ways that seemed somewhat promising in the seventies.⁴⁸ Alas, those ways no longer seem plausible,⁴⁹ not to mention that we no longer seem to have IBS at hand to reduce the semantic to the psychological.

⁴⁶ Chisholm (Op. cit.).

⁴⁷ Avramides (1989) argues that Gricean semantics should be divorced from this motivation.

⁴⁸ See e.g. Loar (1981) and Schiffer (1982).

⁴⁹ See Schiffer (1987).

References

- Austin, J.L. (1962). *How to Do Things with Words*, ed. J. O. Urmson (Oxford University Press).
- Avramides, A. (1989). *Meaning and Mind: An Examination of a Gricean Account of Language* (MIT Press).
- Buchanan, R. (2010). "A Puzzle about Meaning and Communication," *Noûs* 44: 340-71.
- Buchanan, R. and Ostertag, G. (2005). "Has the Problem of Incompleteness Rested on a Mistake?," *Mind* 114: 889-913.
- Bennett, J. (1976). *Linguistic Behaviour* (Cambridge University Press).
- Chisholm, R. (1958). "Chisholm-Sellars Correspondence on Intentionality," in Marres (1958).
- Chomsky, N. (2006). *Language and Mind*, third edition (Cambridge University Press).
- Evans, G. (1973). "The Causal Theory of Names," *Aristotelian Society Supplementary Volume* 47: 187-208.
- Evans, G. and McDowell, J. (eds) (1976). *Truth and Meaning: Essays in Semantics* (Oxford University Press).
- Gilbert, M. (1981). "Game Theory and Convention," *Synthese* 46: 41-93.
- Grice, H.P. (1957). "Meaning," *Philosophical Review* 66: 377-88.
- _____ (1968). "Utterer's Meaning, Sentence-Meaning, and Word-Meaning," *Foundations of Language* 4: 225-42.
- Laurence, S. (1996). "A Chomskian Alternative to Convention Based Semantics," *Mind* 105: 269-301.
- Loar, B. (1976). "Two Theories of Meaning," in Evans and McDowell (1976): 138-61.
- _____. (1981). *Mind and Meaning* (Cambridge University Press).
- Marres, A. (ed). (1958). *Intentionality, Mind, & Language* (University of Illinois Press).
- Neale, S. (1992). "Paul Grice and the Philosophy of Language," *Linguistics & Philosophy* 15: 509-59.
- Pinker, S. (1994). *The Language Instinct* (W. Morrow & Co).
- Schiffer, S. (1972). *Meaning* (Oxford University Press).
- _____ (1978). "The Basis of Reference," ...
- _____ (1981). "Indexicals and the Theory of Reference," *Synthese* 49: 43-100.
- _____ (1982). "Intention-Based Semantics," *Notre Dame Journal of Formal Logic* 23: 119-56.

- _____ (1987). *Remnants of Meaning* (MIT Press).
- _____ (2003). *The Things We Mean* (Oxford University Press).
- _____ (2006). “Two Perspectives on Knowledge of Language,” *Philosophical Issues* 16: 275-87.
- Sawyer, S. (2009). *New Waves in Philosophy of Language* (Palgrave Macmillan).
- Speaks, J. (2009). “Introduction, Transmission, and the Foundations of Meaning,” in Sawyer (2009).
- Strawson, P. (1964). “Intention and Convention in Speech Acts,” *Philosophical Review* 73: 439-60.