Mandy Simons

Carnegie Mellon University, Department of Philosophy

2000 Forbes Ave

Pittsburgh PA 15217

USA

phone: 412.268.5083

fax: 412.268.1440

simons@andrew.cmu.edu

July 29, 2009

103.   Implicature

1. Introduction
2. The Gricean conception of conversational implicature
3. A case study: scalar implicature
4. Embedded implicature
5. Alternate models and competing conceptions
6. Formal approaches

This article reviews in detail Grice's conception of conversational implicature, then surveys the major literature on scalar implicature from early work to the present. Embedded implicature is illustrated, and it is explained why this phenomenon poses a challenge to the Gricean view. Some alternate views of conversational implicature are then presented. The article concludes with a brief look at formal approaches to the study of implicature.

1. Introduction

Conversational implicature is the phenomenon whereby a speaker says one thing and thereby conveys (typically, in addition) something else. For example, in (1) below, Harold *says* that Sally should bring her umbrella, but further conveys that (he believes that) it is likely to rain. This is a standard case of the phenomenon under examination.

(1) Sally: What's the weather going to be like today?
    Harold: You should bring your umbrella.

Conversational implicature was identified and named by the philosopher Paul Grice in his paper *Logic and Conversation*, originally presented at Harvard in 1969. Much of today's linguistic pragmatics has its origins in the insights of that paper, and concerns itself in some fashion with some aspect of conversational implicature.

2. The Gricean conception of conversational implicature

2.1. Implicature as part of what is meant

For Grice, what a speaker *means* by an utterance is the total content which she thereby intends to communicate (see also article 2 *Meaning, Intentionality and Communication* and article 5 *Meaning in Use*). One component of what is *meant* is what is *said*: roughly, the truth conditional content linguistically encoded in the utterance. The remainder – what is meant but not said – is what Grice calls *implicature*. Implicature itself subdivides into two major categories: conventional and conversational. Conventional implicature is content which is conventionally encoded but non-truth-conditional (cf. article 106 *Conventional Implicature*). In this article, we will be concerned with *conversational implicature*: implicatures that arise by virtue of general principles governing linguistic behavior. In "Logic and Conversation" (Grice 1975: henceforward, *L&C*) and "Further Notes on Logic and Conversation" (Grice 1978: hence, *FN*), Grice introduces the phenomenon of conversational implicature and lays out the principles which allow speakers to systematically mean more than they say.

2.2. The Theory of Conversational Implicature

To account for the phenomenon of conversational implicature, Grice proposes that there are certain norms of conversational behavior, norms which are mutually known and typically adhered to by conversational participants. These norms prevent conversation from consisting of "a succession of disconnected remarks," and, at each stage in a conversation, render certain possible conversational contributions "unsuitable" (*L&C* 26). Grice summarizes the effect of these norms as a single overarching principle, which he calls the Cooperative Principle:

> Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged.

This principle has little force without further clarification of what is in fact required of conversational contributions. Grice specifies this further in what he calls *Maxims of Conversation*,

formulated as rules governing allowable conversational moves. Grice organizes these maxims into four categories: Quality, Quantity, Relation and Manner. In current usage, these terms are used to designate the specific maxims Grice proposed. Grice himself, however, gives these specifically as *categories* of maxims "under one or another of which will fall certain more specific maxims and submaxims" (p.26), apparently envisaging the possibility of substantial lists of conversational rules. The maxims which Grice proposes are listed below, in some cases slightly reformulated from the original:

Conversational Maxims

Quality

Supermaxim: Try to make your contribution one that is true

1. Do not say what you believe to be false.
2. Do not say that for which you lack adequate evidence.

Quantity

1. Make your contribution as informative as is required (for the current purposes of the exchange).
2. Do not make your contribution more informative than is required.

Relation

Be relevant

Manner

Supermaxim: Be perspicuous

1. Avoid obscurity of expression
2. Avoid ambiguity
3. Be brief (avoid unnecessary prolixity)
4. Be orderly

The view that conversation is a norm-governed activity provides the basis for Grice's account of how conversational implicatures arise. The general structure of the account is this: There is a standing presumption that speakers produce utterances which are in accord with the Cooperative Principle and its maxims. Interpreters will assign to an utterance an interpretation in accord with this presumption. In some cases, this will require the interpreter to attribute to the speaker the intention to communicate something more than, or different from, what she has actually said. In identifying what the speaker intends, the interpreter will rely on three things: first, her observation about what the speaker *said* (i.e. the truth conditional content expressed) and the form in which it was expressed; second, the presumption of cooperativity; and third, any world knowledge that might be relevant. Speakers can anticipate this behavior of interpreters, and thus can predict that particular utterances will be understood as conveying something more than or different from what is literally said. The fact that it is common knowledge that the CP is in effect thus allows speakers to implicate, and interpreters to identify implicatures.

Grice characterizes conversational implicature in the following way (slightly simplified from the original):

> A man who, by saying ... that *p* has implicated that *q*, may be said to have conversationally implicated that *q*, provided that:
>
> > 1. he is to be presumed to be observing the conversational maxims or at least the Cooperative Principle
> >
> > 2. the supposition that he thinks that *q* is required in order to make his saying *p* (or doing so in *those* terms) consistent with this presumption and
> >
> > 3. the speaker thinks (and would expect the hearer to think that the speaker thinks) that it is within the competence of the hearer to work out, or grasp intuitively, that the supposition mentioned in 2. is required.

Grice's presumption here is that such suppositions of the speaker, when recognized by the hearer, will be understood to be meant.

Clause 2 of this definition is quite problematic, as it makes it a condition on implicature that the implicature (the "required supposition") be uniquely adequate to maintain the presumption of cooperativity. This is typically too strong: in almost any case of conversational implicature, there are

multiple candidate suppositions, any of which might render the utterance cooperative. This point is noted by Grice (see the final paragraph of *L&C*). Davis (1998) considers this (along with some other difficulties) to completely undermine the Gricean construction. The issue is worth further exploration, but I will not attempt it here.

Let's now make things clearer by examining some specific cases. Consider again the sample discourse from above, repeated here as (2):

(2) Sally: What's the weather going to be like today?

   Harold: You should bring your umbrella.

Observation: Sally is likely to conclude that Harold means to inform her that it is likely to rain. How so? First, Sally presumes that Harold is speaking in accord with the CP. Among other things, this means that she presumes that he intends his answer to be relevant. Now, strictly speaking, Harold's instruction is not an answer to the question Sally has asked: it says nothing about the weather. But *because* of her presumption, Sally presumes that what Harold says *is* relevant in some way to her question. It immediately occurs to her that one uses an umbrella when it rains; and that Harold is likely to tell her to bring an umbrella if he believes that it is going to rain. If she attributes this belief to Harold, and assumes that he intends, via his utterance, to communicate this belief to her, then she has successfully interpreted his utterance in a way which renders his behavior consistent with her presumption of his cooperativity. As Harold can, moreover, plausibly assume that Sally will reason in this way, he implicates that (he believes that) it is going to rain.

Note that this case illustrates the failure of the uniqueness condition on implicature (Clause 2 of the definition above). Suppose it is common knowledge between Sally and Harold that Sally uses her umbrella as a sunshade when the weather is hot. Then Sally might just as well have attributed to Harold the belief that it would be hot and sunny, and take him to intend to communicate this. So there are (at least) two candidate suppositions that would render Harold's utterance cooperative. On the other hand, contextual factors (such as the interlocutors' common knowledge of recent weather) might well render one of these suppositions far more likely or reasonable. This line of thought might offer a resolution of the difficulty.

The example in (2) involves a Relevance implicature. Implicatures can be generated via any of the maxims (or combinations thereof). Here is one which relies on the first Maxim of Quantity:

(3) Harold: Which of Bobby's teachers did you talk to at the picnic?

    Sally: Mrs. Smith and Mr. Jones.

Here, Sally implicates that Mrs. Smith and Mr. Jones were the *only* teachers that she talked to. This is by virtue of the first maxim of quantity. Given the assumption that Sally is abiding by this maxim, Harold must assume that she will provide all the information relevant to his question. If (she believed that) she had talked to additional teachers, then it would constitute a violation of the maxim to fail to mention them. So, to maintain the premise that Sally *is* abiding by the maxim, Harold must assume that Sally (believes that she) spoke to no other teachers. As Sally, moreover, can assume that Harold will recognize the required assumption, she implicates that she talked to no other of Bobby's teachers.

2.2.1. Characteristics of conversational implicature

In the final pages of *L&C*, Grice identifies certain characteristic features of conversational implicatures. The central ones are these:

1. Calculability: if some element of content is a conversational implicature, then it should be possible to provide an account of how it is calculated on the basis of what is said plus the maxims.

2. Nondetachability: On Grice's view, implicatures other than Manner implicatures are calculated on the basis of what is said – roughly, on the basis of the truth conditional content expressed. Hence, other ways of expressing the same truth conditional content in the given context should give rise to the same implicature. That is, implicatures are *nondetachable* from a particular truth conditional content.

3. Cancelability (of generalized conversational implicature): Because conversational implicatures are not part of the encoded or conventional content of any linguistic item, and because their presence is dependent on (more or less) specific assumptions, including

the assumption of the cooperativity of the speaker, then it should be possible for an expected implicature to be contextually canceled.

These features, particularly cancelability, are sometimes taken as diagnostics of conversational implicature. However, Grice did not intend them this way, as he clarifies in *FN* (p.43). Sadock (1976) provides thorough arguments showing that none of these features, either separately or together, can serve as robust diagnostics of conversational implicature, as none are either necessary or sufficient conditions. The arguments are too lengthy to rehearse here in detail, but a couple of points are worth mentioning. With respect to calculability, Grice and Sadock agree that it is not sufficient to establish the presence of a conversational implicature, because what starts life as a conversational implicature may become conventionalized. Nondetachability is neither a necessary nor a sufficient condition. It is not necessary, because it is not a feature of Manner implicatures. It is not sufficient for the identification of conversational implicatures, because it is also a feature of entailments.

Sadock gives a second argument against the necessity of nondetachability, observing that there are cases of truth conditionally equivalent sentences whose utterance does not give rise to the same implicatures. Consider sentences (4) and (5) as answers to the question: *Did you eat any of the cake?*

(4) I ate some of the cake.

(5) I ate some and possibly all of the cake.

Obviously, (4) implicates that the speaker did not eat all of the cake, while (5) just as obviously does not so implicate. Yet the two sentences (by assumption) have the same truth conditional content i.e. in both cases the same thing is said. Hence, the implicature is *not* nondetachable: utterances of truth conditionally identical S and S' do not both produce the implicature.

One possible response to this argument is simply that the definition of nondetachability requires refinement: it should exclude candidates which are truth conditionally equivalent to the original but include an explicit denial of the potential implicature. Other reformulations of (4) do preserve the implicature. Consider *I ate part of the cake* or *I ate a bit of the cake*.

A second response is that the input to conversational implicature calculation is not simple truth conditional content, but some more structured entity. Arguments for this position are given by Gazdar (1979) and Atlas & Levinson (1981).

Finally, we turn to cancelability. First, note that the type of cancelability Grice has in mind involves the speaker being explicit that she is opting out of the observation of the CP, or the context being one which makes clear that the speaker is opting out. In *FN*, he gives the example of a speaker who is giving a clue in the course of a treasure hunt saying:

(6)   The prize is either in the garden or the attic, but I'm not telling you which.

In this context, the typical implication from a disjunction, that the speaker does not know which disjunct is true, is suppressed.

Sadock discusses a different type of cancelation, where the speaker explicitly denies the usual implicature, as in:

(7)   Some philosophers are crabby, and I don't mean to say that some aren't.

In the current literature, when people discuss implicature cancelation, the latter is usually what is intended.

Grice seems to consider that cancelability can only apply to *generalized* conversational implicatures. What he seems to have in mind is that we make observations about what is normally or typically implicated by the use of a particular expression, and compare it with what (if anything) is actually implicated by the use of that expression in some specific situation. We clearly cannot make the same sort of comparison in the case of *particularized* implicatures. For example, no-one would claim that the sentence *I have to cook dinner* normally or typically implicates *I am not going to read you a story*, but certainly an utterance of that sentence might well so implicate if I say it in response to my six year old's request in easily imaginable circumstances. Nonetheless, we sometimes find cases like these:

(8) C: Mommy, will you read to me?

  M: I have to cook dinner. So if I read to you now, will you play by yourself for a while afterwards, so I can get dinner done?

The first sentence, if uttered alone in this context, might well be used to implicate "no." The entire string, however, makes clear that this is intended to launch a "yes, but..." response. So, there is some temptation to say that the second sentence cancels the implicature arising from the first. This is similar to a second way of understanding implicature cancelation in the generalized case. In cases like (12), one might say that the use of the first clause *does* generate the implicature, but that the implicature is canceled – that is, the initial clause is reinterpreted – in light of the content of the second.

Which way we should see it depends in part on our assumptions about when implicature calculation takes place. It is clear that Grice assumes throughout most of his writing on the subject that the input is at least a complete proposition. The examples used typically involve a single sentence generating an implicature. But it is perfectly consistent with the Gricean model that the semantic content of a multi-sentence conversational contribution – presumed to be several propositions – could be the basis for a process of implicature calculation. If implicatures are calculated in this way, example (12) could only be said to involve cancelation in the sense that an implicature that typically arises fails to do so; and in the case of (13) it would not be sensible to talk about cancelation at all.

Cancelability remains an important diagnostic for distinguishing between conventional content and inferred content (although see again Sadock's arguments concerning ambiguity). However, it is important, in making use of this notion, to be clear just what we mean by it in any particular case.

2.2.2. Subtypes of conversational implicature

The Quantity implicature in (3) above straightforwardly fits Grice's own characterization of conversational implicature. But the Relevance implicature in (2) fits it rather awkwardly. Although I formulated it this way above, it is somewhat odd to say that Sally recognizes that Harold *presumes* that it will rain, and therefore takes him to intend to communicate this. It seems more natural to say that Sally recognizes that Harold *presumes* that she would want to have her umbrella with her if it

rains, and thus infers, from his recommendation that she bring her umbrella, his intention to communicate that it might rain. Let's call the identified presumption a *background implicature*, and the communicated proposition, that it might rain, a *foreground implicature*.

Sally's recognition of the background implicature seems to make use of standard Gricean reasoning: searching for a way to interpret Harold's utterance as cooperative, she looks for a presumption he might be making which would render what he said relevant to her question. However, the background implicature is nonetheless not a true implicature in Grice's sense; for recall that for Grice, implicature is a subcategory of speaker meaning; and what a speaker means is what he intends to communicate. In the kind of conversation we are imagining between Sally and Harold, it would not typically be Harold's intention to communicate that Sally likes to have her umbrella with her when it rains.

Here we reach the first of many choice points, where we will have to decide: is our goal to follow Grice's conception as closely as possible? Or to use his proposal as a jumping off point from which to develop an empirically adequate and explanatory theory of pragmatics? For linguists, the answer is typically the latter. As a first departure from Grice, we might propose using the term *conversational implicature* for any inference the speaker intends the addressee to make on the basis of the assumption that he is being cooperative. The distinction made above between background and foreground implicature can be further explicated by distinguishing between implicatures which are not meant (in Grice's sense) and those that are. (For further discussion, and identification of background implicatures with presuppositions, see Simons 2004, 2008.)

One additional Gricean distinction remains to be made: that between *particularized* and *generalized* conversational implicature. Example (2) involves a particularized implicature. Here the implicature relies to a high degree on the conversational context and on specific background assumptions attributed to the interlocutors. Changes in the context or in these assumptions easily change the implicature or eliminate it. For example, if Harold had uttered the same sentence in a conversation in which the interlocutors we trying to figure out how to knock a ball out of a tree, it would give rise to no implicatures at all about the weather. There is thus "no room for the idea" (to use Grice's phrase) that the implicature is associated with this particular form of words or with the expression of this content.

In contrast, in the case of Generalized Conversational Implicatures, exactly this claim is warranted. GCIs are implicatures which normally arise when a particular form of words is used. The only example Grice gives of a GCI in *L&C* involves the interpretation of indefinite NPs, which turns out to be a rather complex case. In *FN*, he discusses the implicature associated with the use of *or* that the speaker does not know which disjunct is true. Grice notes that "noncontroversial examples [of GCIs] are perhaps hard to find, since it is all too easy to treat a generalized conversational implicature as if it were a conventional implicature" (*L&C*, p.37).

In the post-Gricean literature, an element of content is identified as a GCI if, on the one hand, it can be explained as a Gricean implicature and is cancelable; but, on the other hand, its occurrence is not dependent on irregular features of context, but only on the basic assumption that the speaker is speaking in accord with the CP. Some authors deny that there is any theoretically significant distinction between generalized and particularized implicatures (e.g. Hirschberg 1991); while others hold the opposite view (see Levinson 2000, discussed below in section 5).

Presuppositions are also considered by some authors to be a subtype of implicature, although this view is far from widespread. The bulk of linguistic work on presupposition since the 1970's has focused on the projection problem: assuming the association of presuppositions with particular atomic sentences, how do we account properly for the presuppositions of complex sentences in which the atomic sentences are embedded? The possible relationship between implicature and presupposition becomes salient when we focus on a different question, namely, how do presuppositions arise in the first place? Many authors have suggested that some or all presuppositions arise via the workings of Gricean conversational principles: see among others Stalnaker 1974, Kempson 1975, Wilson 1975, Boër & Lycan 1976, Atlas 1977, 1978, 2005, Atlas & Levinson 1981, Grice 1981, Simons 2001, 2004. The idea underlying all of these proposals is formulated succinctly by Stalnaker: "one can explain many presupposition constraints in terms of general conversational rules without buildijng anything about presuppositions into the meanings of particular words or constructions" (1974: 212). For a discussion of this issue in the broader context of the literature on presupposition, see Article 102.

22.3. Rational underpinnings of the Maxims

After laying out the CP and the maxims in *L&C*, Grice raises "a fundamental question about the Cooperative Principle and its attendant maxims, namely, what the basis is for the assumption which we seem to make, that talkers will in general … proceed in the manner that these principles prescribe" (p.28). He goes on to say: " I would like to be able to think of the standard type of conversational practice … as something that it is *reasonable* for us to follow," and sets out the following as a set of claims to be explored:

1. There are certain goals that are fundamental to conversational exchange.
2. These goals can be accomplished only in conversational exchanges conducted in accordance with the CP.·
3. Therefore rational speakers will behave in accord with the CP.

Claim 2 in this list seems unnecessarily strong. We might do better with 2':

2'. These goals can be best accomplished (i.e. most efficiently, with least effort of both speaker and hearer) in conversational exchanges conducted in accordance with the CP.

As part of the effort to establish these rationalist foundations for the account, Grice demonstrates that each of the Maxims has an analog in the non-linguistic domain, in other types of goal-oriented interactions. The effort, however, does not go much beyond this, and Grice leaves as an open question whether his conversational principles can be independently motivated. But this proposal sets a particular foundational project for pragmatics, namely, to determine the rational underpinnings of pragmatic principles.

Perhaps because of the label *Cooperative Principle*, it is standard to suppose that observation of the CP and maxims is tied to a certain amount of explicit cooperativity between the discourse participants. What is clearly not required is alignment of practical goals: Grice was clear that the CP must be construed as in effect even in cases like quarreling. The degree of cooperativity envisioned by Grice seems to go no further than a mutual desire on the part of the interlocutors to use language in a way that facilitates communication between them.

Nonetheless, there are some difficult cases: what account is to be given of apparent implicatures in cases where the misaligned goals of interlocutors should lead them to expect their discourse partners to be, for example, intentionally underinformative? (For discussion within a game theoretic framework, see Franke, de Jager & van Rooij 2008.) Both empirical and theoretical questions concerning such issues remain open.

2.4. A friendly amendment: Horn's Q- and R- principles

As noted, Grice seems to envision a long, open ended list of specific submaxims falling under each of his four categories. Horn (1984) develops Grice's view in a different direction. Setting aside Quality as having a special status, Horn proposes a reduction of the remaining Gricean principles to just two: Q ("to evoke Quantity") and R (Relation).

Horn's Principles

(Q)   Make your contribution sufficient; say as much as you can (given R).

(R)   Make your contribution necessary; say no more than you must (given Q).

The Q-principle corresponds to the first maxim of Quantity. R is intended to encompass Relation, the second maxim of Quantity, and Manner.

The precise wording of the principles is perhaps less important than the underlying idea, which Horn attributes to Zipf (1949) and Martinet (1962). The idea is that there are two basic and conflicting forces at work in the linguistic domain, representing the interests of the speaker and the distinct interests of the hearer. It is in the interest of the hearer/interpreter to receive maximum (relevant) information and to receive it in the clearest formulation available. The Q-principle encapsulates this as a hearer-oriented requirement imposed on the speaker. However, it is in the interest of the speaker to minimise the effort involved in producing her message; this interest underlies the R principle.

Horn emphasizes two points: first, that his two principles together do the work of Grice's categories of Quantity, Relation and Manner; second, that the two principles are in conflict. This latter observation provides the basis of an explanation of certain cases of miscommunication, where

different parties may be exploiting or emphasizing different principles. But it also raises an empirical question, namely, in cases of successful communication involving implicature, on what grounds do speakers/hearers determine which principle takes precedence? Consider for example the following cases discussed by Horn:

(9) I broke a finger last week.

(10)   I slept on a boat last night.

At issue is the natural interpretation of the indefinite. We are inclined to understand the speaker of (14) as saying that she broke one of her own fingers. This would be, in Horn's view, an application of the R-principle: what is meant is something informationally richer than what is said; the speaker has provided the minimum information necessary to convey the total intended message. In contrast, it is natural to understand the speaker of (15) as saying that she slept on a boat not belonging to her. This, Horn takes to involve an application of the Q-principle: as the speaker has provided no more specific information about the boat, we take it that no additional information is available or relevant. (It is not entirely transparent how the principles are to apply in each case to produce the specific interpretations. What is important here is that interpretation goes in two different directions with two superficially similar sentences.)

Horn's brief discussion suggests that there may be many different, and differently weighted, considerations involved in resolving a maxim-clash. The considerations may be specific to the construction involved, and hence it is likely that there is no global answer to the question of which maxim or principle applies; the question may well require a case-by-case answer.

The problem of clashes between maxims had received prior attention in Atlas & Levinson (1981). They offer a Principle of Informativeness to account for example (14) and a host of additional cases where the speaker is understood as conveying something more specific than what is literally said, noting that in many cases, a conflict exists between this principle and Grice's maxim of Quantity.

3. A case study: scalar implicature

So far, we have reviewed Grice's foundational proposal and some proposed extensions. In this section, we look in detail at one particular linguistic phenomenon – scalar implicature – which has commonly been analyzed in Gricean terms.

3.1. The basic phenomenon and the Gricean account

The following are standard examples of scalar implicature. In each case, the b. sentence represents an implication that would normally arise if the a. sentence was uttered in ordinary discourse:

(11)   a.   Some of my friends are Jewish. ↝

         b.   It is not the case that more than some of my friends are Jewish.

(12)   a.   Their new house is large.   ↝

         b.   Their new house is not enormous.

(13)   a.   It's possible that I'll get funding for next year. ↝

         b.   It's not probable that I'll get funding for next year.

That the content of the b. sentences is an implication and not an entailment can be demonstrated by the possibility of cancelation, as in:

(14)   Some of my friends are Jewish; in fact, most of them are.

(15)   Their new house is large; indeed, it's enormous.

We will assume for now without argument that these implications are implicatures.

The implicatures in question are clearly generalized, not particularized. This is evident, first, from the fact that we can identify them in the absence of any particular information about a context of utterance. Second, the implicatures seem to arise by virtue of the presence of particular expressions, or at least, expressions with particular content: *some*, *large* and *possible.* It does indeed seem that it is the content that is relevant and not the particular lexical item: the same implicatures would arise from utterances of the sentences below:

(16)   A few of my friends are Jewish.

(17)   Their new house is big.

(18)   I might get funding for next year.

These implicatures are amenable to what seems a very straightforward Gricean explanation, based on the first Maxim of Quantity. Recall that this Maxim says:

Make your contribution as informative as is required (for the current purposes of the exchange).

It is crucial for the analysis we are about to give to remember that Quantity is circumscribed not only by Relevance but also by Quality: while you should give all the information that is required, you should only give reliable information.

Now, consider a speaker who utters (16)a. Assume that it is relevant to the current conversation roughly what proportion of her friends are Jewish. Then Quantity1 enjoins her to be fully informative with respect to this question.

One natural way (although by no means the only way) to quantify informativity is in terms of entailment: If p entails q, then p is more informative than q. Now consider again the speaker who utters (16)a., and assume that she is being fully informative in the sense just described. With respect to the (presumed) question at hand, her utterance rules out the possibility that none of her friends are Jewish, but leaves open the possibility that some, many, most or all of her friends are. She could have ruled out these possibilities by making a more informative statement, e.g. by saying:

(19)   Many/most/all of my friends are Jewish.

But we presume that the speaker has made the most informative utterance she can given the requirements of Quality. So this leads to the conclusion that she is not in a position to utter any of the variants of (24). But she is presumably in a position to know (more or less) what proportion of her friends are Jewish. Therefore, we conclude further that she is not in a position to make the stronger assertions because they are not true. That is, we conclude that (16)a. was the most informative true sentence the speaker could utter relative to the question at hand, and thus conclude from the utterance that some of the speaker's friends are Jewish, and no more than some are.

Parallel derivations can be given for the implicatures in (17) and (18). In each case, the implicature is the negation of a sentence (proposition) which entails (i.e. is informationally stronger than) the actual utterance. The negated proposition in each case represents a stronger alternative which (in some sense) the speaker "could have said" instead of what she actually said.

3.2. The scalar solution

In the account just sketched, I carefully avoided any reference to the presence of particular expressions. The account makes reference only to propositions ordered by entailment. But it is not really feasible to maintain so pure an account. Reference to expressions seems necessary to identify the alternative "possible utterances" relative to which these implicatures are calculated. For there are of course all sorts of other, stronger utterances which the speaker could have made whose content is potentially relevant to the same conversation. The speaker of (16)a. could have said:

(20)   Some of my friends are Jewish and they eat kosher.
(21)   Some of my friends are Jewish and some of my friends are Muslim.

or any of an infinity of other possibilities.

In fact, there are imaginable conversational contexts in which utterance of (16)a. *would* implicate the negation of some such other stronger sentence. For example, suppose that (16)a. were uttered in response to the question:

(22)   Do you know anyone who isn't a Christian?

In this context, utterance of (16)a. would plausibly give rise to the implicature that the speaker has no friends who are neither Christian nor Jewish e.g. has no Muslim friends; and this could be explained by reference to the speaker's failure to utter this particular stronger alternative to her actual utterance. So it is not that random stronger alternatives do not give rise to implicatures based on the same kind of reasoning. Rather, they only do so when licensed by special features of the context, that is, they are particularized implicatures. What this tells us is that to fully explain the cases in (16)-(18), we need to say what allows for a context *independent* judgment about the relevant alternatives.

The solution offered by Horn (1972), now the standard account, invokes the notion of a scalar term. These are terms which cluster with other semantically related terms on a scale of informational strength, with relative informativity being measured, as above, in terms of entailment relations. The standard formalization of such Horn scales comes from Gazdar (1979): If Q is a Horn scale of form $\langle \alpha_1, \alpha_2, \alpha_3 ... \alpha_n \rangle$, then each element in Q is logically weaker than any element to its left in the following sense: given a pair of sentences S, S' identical except that where S contains α_i, S' contains α_j and i>j, then, if α_i/α_j are not embedded under any operator, S entails S'. Let's call the relation between elements in such a scale *asymmetric lexical entailment*.

The basic Horn/Gazdar account of generalized scalar implicatures then runs like this: Utterance of a sentence containing an element in a Horn scale renders salient the set of stronger scalar alternatives to the original utterance (i.e. for the unembedded case, sentences identical to the original but with the scalar term replaced by a scale-mate to its left in the scale), regardless of the particular conversational context. A Quantity-based inference generates scalar implicatures based on these alternatives.

We'll see shortly that entailment-based scales do not give us a complete account of scalar implicature. But first, let's review two questions that arise even given this fairly simple model. First

question: how many implicatures do scales generate? For example, given the scale <*all, most, many, some*>, does utterance of sentence S containing *some* implicate the negations of all of the scalar alternatives to S based on this scale? Does it implicate just the negation of the next strongest, which in turn entails the negations of the remainder? Does it implicate just the strongest? Hirschberg 1985 argues that the answer may depend on the context. Consider for example the pair of discourses below:

(23) a. A: Are all mushrooms poisonous?

         B: Some are.

      b. A: Are many mushrooms poisonous?

         B: Some are.

Each question makes salient a different sub-part of the relevant scale, changing the implicature generated.

A second question raised by scalar implicatures concerns what Hirschberg calls their epistemic force. (In fact this question arises for all cases of Quantity implicature, but has been discussed in particular in the context of scalar implicatures.). This issue is discussed in almost all of the early research on scalar implicature (Horn 1972, 1989, Gazdar 1979, Hirschberg 1985; see also Soames 1982: 455-6, Levinson 1983: 135, Groenendijk & Stokhof 1984, Sauerland 2004). Soames, Horn (1989) and Sauerland each adopt more or less the following position: Consider again example (16)a. above. The Maxim of Quantity, bounded as it is by the maxim of Quality, supports only the weak inference that the speaker *does not know* any stronger assertion to be true. However, in some cases, like that of (16)a., we can assume that the speaker has complete knowledge of the situation she is describing. In such cases, the weak inference supported by the Maxim can be strengthened: the interpreter can infer that the speaker *knows* the stronger alternative *not* to be true. In our example, the interpreter can infer that the speaker knows that it is not the case that many, most or all of her friends are Jewish. Then, by the factivity of knowledge, it is inferrable that it is not the case that many, most or all of her friends are Jewish.

This Gricean predication does not, though, quite match up to intuition. Weak scalar implicatures are rather unusual, typically arising only when the speaker has explicitly acknowledged the incompleteness of her information, as in:

(24)   I've only looked at a few pages of the manuscript, but I can already tell you that some of the pages require corrections.

It is much more typical for scalar implicatures to be epistemically unmodified, even in circumstances where, on reflection, the interpreter might have less than complete confidence in the speaker's authority or evidence. (For more detailed discussion of this topic, see Sauerland 2004 and van Rooij & Schultz 2004.)

Now we come to two questions much more central to the notion of scale. The first: what precisely are the elements of the scale? The discussion so far suggests that scales consist of lexical items. But Gazdar (1979) proposes that the elements should be semantic values or expressions in a formal language used to represent the semantic content of sentences. Gazdar argues for this on the basis of simplicity: For one thing, a language might contain multiple lexical items with similar content (e.g. *large* and *big*) which occupy the same location in a scale (e.g. relative to *huge*).

The second core question raised by scales themselves is this: Where do scales come from? Or where do they reside? The ideal answer (from a Gricean perspective) would be that scales are merely a formal construct capturing relations between lexical items (or semantic values of items) which are fully reconstructible just on the basis of lexical content. That is, we do not need to stipulate that, for example, <*all, most, many, some*> is a scale, and a language user does not need to learn that it is. Any speaker who knows the meanings of these terms will recognize that they form a scale by virtue of the lexical entailment relations which hold between them.

But additional observations have led many to conclude that such an answer is not available. The additional observation is that not every entailment scale is a Horn scale i.e. a scale supporting scalar implicature. Gazdar (1979) gives a simple example to illustrate this: *regret* asymmetrically entails *know*, but use of *know* does not implicate *not regret*. The example shows that asymmetric lexical entailment is not a sufficient property to qualify a pair of words (or their meanings) as Horn scale-mates.

There is a second type of case where asymmetric lexical entailment is insufficient for Horn scale membership, giving rise to what has recently been dubbed the Symmetry Problem. Consider the pair: *some but not all, some*. The first asymmetrically entails the latter; but clearly, an utterance of (30)a. does not implicate (30)b.:

(25)   a.   I ate some of the cookies.

         b.   It is not the case that I ate some but not all of the cookies

for the conjunction of a. and b. is equivalent to "I ate all of the cookies" – and this is the negation of what (30)a. is actually observed to implicate!

Horn (1972) addresses problem cases of this sort by positing a requirement that elements in scales have the same degree of lexicalization. As English contains no single word which means "some but not all," the semantic content of this expression cannot be a scale-mate with the content of *some*. This constraint can perhaps be given a Gricean motivation in terms of the Maxim of Manner, the idea being that the alternatives to be considered should only be those which do not increase the complexity of the utterance. (See also Matsumoto 1995, Katzir 2007.)

In light of these difficulties, many researchers adopt the position that Horn scales are linguistically given. So, although we can make certain general observations about the relations which hold between members of a Horn scale, we cannot predict merely on the basis of the linguistic properties of an item whether or not it will occur on a scale, or what its scale-mates will be. It is simply a contingent fact about a language what scales it contains, and what the members of those scales are.

3.3. Beyond scales

So, we reach the following position: We posit Horn scales as linguistic elements. Because of the existence of these scales, an utterance U of sentence S containing a scalar term makes salient a set of alternative sentences. We can then apply basic Gricean reasoning utilizing Quantity1 to give an account of scalar implicatures.

But that basic Gricean reasoning relied on cashing out informativity in terms of asymmetric entailment. Yet there appear to be cases of scalar implicature which don't rely on scales ordered by this relation. Horn (1989) observes several cases where lexical items within a particular domain are intuitively rankable (to borrow a term from Hirschberg 1985), but are not related by lexical entailment. Horn's examples include military rank (*private, corporal, sergeant...*), life stages (*newborn, infant, toddler, preschooler...*), and legal violations (*tort, misdemeanor, felony, capital crime*). These, Horn suggests, give rise to implicatures following the same pattern as entailment based scalar implicatures, as illustrated by the examples below:

(26)   A: Do they have any pre-schoolers in the program?

    B: They have toddlers. [~ toddlers but no older pre-schoolers]

(27)   A: Have you worked on any capital cases?

    B: I've worked on several felony cases. [~ felony cases but no capital cases]

Hirschberg (1985) goes further in her observations. She points out that scalar implicature-like phenomena arise with a variety of non-linear rankings including whole/part relationships, type/subtype, and prerequisite orderings. Examples of implicatures based on these relations (all borrowed from Hirschberg Ch.5) are given below:

(28)   A: Did you read that section I gave you?
    B: I read the first few pages.

(29)   A: Do you have a pet?
    B: We have a turtle.

(30)   A: Is she married?
    B: She's engaged.

Similarly, scalar implicatures can be produced where context makes salient some set of alternatives. Here, affirmation of one alternative gives rise to an implicature that no other alternative can be affirmed. So consider:

(31)   A: Do you have apple juice?

   B: I have grape or tomato or bloody mary mix.

On the basis of such observations, Hirschberg (p.125) concludes that scalar implicatures are supported by the class of partially ordered sets i.e. any relation which is reflexive, antisymmetric and transitive. Any set of lexical items referring to entities or processes which stand in such a relation may provide the basis for a scalar implicature.

In order for a scalar implicature to be generated, the ordering which underlies it must be somehow salient. But context can render all kinds of ad-hoc orderings salient, given adequate background. If the Hirschberg-type examples really are cases of scalar implicature, then an account of this phenomenon in terms of conventional, linguistically pre-given scales seems highly implausible. On the other hand, one might argue that these examples are simply cases of particularized quantity implicatures, whose explanation requires reference to the conversational context and goals.

What makes it unlikely that the Hirschberg-type cases are unrelated to standard cases of generalized scalar implicature is that both cases clearly involve the construction of alternative possible utterances. It seems likely that an account of the construction or identification of the alternatives in these context dependent cases will be extendable to the classic scalar cases.

A promising candidate for a more general account comes from work on exhaustivity, originating in Groenendijk & Stokhof (1984). These authors were concerned with the derivation of exhaustive answers to questions, as in the following case:

(32)   A: Which students were late today?

   B: Francine was late.

It is natural to understand the question as requiring, and the response as providing, an exhaustive answer. That is, the response is taken as indicating that, of the relevant set of students, *only* Francine was late (for the event in question). The example is much like many of Hirschberg's, which typically involve question/answer pairs, and could similarly be accounted for on the basis of Quantity. Groenendijk & Stokhof, however, take exhaustification to be a semantic effect, the result of applying to the semantic content of the answer an exhaustivization operator, roughly equivalent to *only* in meaning. This account nonetheless requires, like Hirschberg's, positing a salient set of students who are potential alternatives to Francine.

G&S's proposal is extended and applied directly to the scalar case in van Rooij & Schulz (2004) and Schulz & van Rooij (2006). The latter paper offers several revisions to G&S's original formalization of the semantics of exhaustivization, extending the scope of the account and improving on some of the original predictions. More importantly for our interests, Schulz & van Rooij attempt to derive exhaustivization explicitly as a Gricean effect, and explicitly connect exhaustivity to scalar implicature, suggesting (p.245) that many scalar implicatures simply involve exhaustive interpretation. By extension, this suggests that the notion of scale may be dispensible for a large class of cases.

For Groenendijk & Stokhof, exhaustivization is a semantic effect. For Schulz & van Rooij, it is a pragmatic one. Fox (2007) argues for exhaustivization as a syntactic effect, resulting from the interpretation of a covert syntactic element. In terms of its interpretation, Fox's *exh* operator is essentially the same as G&S's, although its occurrence is much less constrained than in their theory. Fox uses his operator specifically to derive scalar implicature effects, unmooring these effects from their Gricean foundations and transplanting them into the grammatical system.

Fox's proposal is one of several in which scalar implicature is derived as a grammatical effect. Landman (1998, 2000) and Chierchia (2004, 2006) propose treatments in which scalar implicature is calculated by the compositional semantic component of the grammar, while remaining sensitive to general contextual factors in some ways. The primary motivation for this "grammatical turn" in the treatment of scalars comes from the phenomenon of embedded implicature, to be discussed briefly in the next section; but is further motivated by the apparent robustness of these implicatures. (But see Geurts 2009 and work cited therein for a challenge to the robustness claim.) What is of particular

interest from a foundational perspective is that this one phenomenon is amenable to analysis in such widely differing ways, facing us with a paradox observed by Grice in *Further Notes*:

> If we, as speakers, have the requisite knowledge of the conventional meaning of sentences we employ to implicate, when uttering them, something the implication of which depends on the conventional meaning in question, how can we, as theorists, have difficulty with respect to just those cases in deciding where conventional meaning ends and implicature begins?

4. Embedded Implicature

Cohen (1971) is one of the earliest published critiques of Grice's theory of conversational implicature. Cohen raises a number of problems for Grice, but his paper is best known for this pair of examples:

(33)   The old king has died of a heart attack and a republic has been declared.

(34)   If the old king has died of a heart attack and a republic has been declared, then Tom will be quite content.

Sentence (38) implies that the king died first, and the republic was subsequently declared. Cohen points out (p.58) that this ordering implication is present also when the conjunction occurs in the antecedent of the conditional in (39). In fact, this implication seems to be part of the *content* of the antecedent; for it is natural to judge the conditional true even if we believe that Tom would be unhappy if the republic had been declared first, and the king had subsequently died. But this implication, according to Grice, is supposed to be a conversational implicature. This thus appears to be a case where a purported conversational implicature falls under the scope of a semantic operator, a phenomenon now dubbed *embedded implicature*. (See also article 98 *Grammatical View of Scalar Implicatures*.)

Since Cohen's work, many additional types of embedded implicature have been identified. So why exactly are these examples problematic for Grice? They are problematic for two reasons, which I call *the calculation problem* and *the compositionality problem* (see Simons to appear).

The calculation problem is this: Grice's account of the calculation of conversational implicatures does not provide a means for the calculation of implicatures from non-asserted clauses. Anscrombe & Ducrot (cited by Recanati 2003) provide a succinct argument to this conclusion:

(a) Conversational implicatures are pragmatic consequences of an act of saying something.

(b) An act of saying something can be performed only by means of a complete utterance, not by means of an unasserted clause such as a disjunct or the antecedent of a conditional.

(c) Hence, no implicature can be generated at the level of an unasserted clause.

Now, perhaps it is to be expected that exceptions will be found for the case of Manner implicatures, which involve attention to the form in which something has been said. But apparent implicatures attaching to embedded clauses are of all kinds. Consider the examples in (40) and (41) below, both involving scalar implicatures:

(35)   Either Kai ate the broccoli or he ate some of the peas. (Sauerland 2004)

    ~ Either Kai ate the broccoli or he ate **some but not all** of the peas.

(36)   Bill thinks that Harriet wants to read some of his papers. (Cf. Chierchia 2004)

    ~ Bill thinks that Harriet wants to read **some but not all** of his papers.

In neither case can the observed implicatures be derived via standard Gricean reasoning from the entire asserted content. (For details, see the cited papers). Suppose then that we want to allow the implicatures to be generated *in situ*. The standard Gricean machinery simply does not provide a means to do this, as the clauses which give rise to the implicatures are not themselves said. The calculation problem, then, is the problem that the Gricean picture provides no means for calculating

implicatures on the basis of content which is not said; but intuitively, syntactically embedded clauses do sometimes seem to give rise to implicatures in their own right.

The second problem, the compositionality problem, can be characterized in two different ways, depending on whether one is interested solely in the consequences of embedded implicatures for Grice's views, or for the theory of interpretation more generally. (For a more general discussion of compositionality, see article 6 *Compositionality*.) Starting with the Gricean perspective, the compositionality problem is this: According to Grice, what is said is supposed to be determined by conventional content (in addition to reference fixing and disambiguation). But in examples like (39)-(41), the implicatures apparently generated by embedded clauses seem to fall under the scope of the embedding operators, and thus to contribute to the truth conditional content expressed: that is, to what is said.But the compositionality problem is not merely a problem for Grice's conception of what is said. It is more broadly a problem for standard models of the interaction between conventionally encoded content and inferentially derived content. Translated into the language of current semantic theory, Grice's model of what we now call the semantics/pragmatics interface tells us that processes of semantic composition are independent of (and are analytically prior to) processes of pragmatic inferencing. But cases of embedded implicature suggest that the correct model is one in which, for example, a linguistic operator can apply to a proposition consisting of both encoded and inferentially derived content.

It is for this reason that local pragmatic effects are viewed by non-Gricean pragmaticists as being of central importance (see Sperber & Wilson 1986, Bach 1994, Recanati 1989, 2004, Levinson 2000) . These effects do not merely require a rethinking of the mechanism whereby pragmatic effects are derived, but of the entire model of interpretation. The authors just mentioned all hold that such effects – what Levinson (2000) calls *pragmatic intrusion* – are ubiquitous, going well beyond the types of examples illustrated here. For such theorists, a central goal of any theory of pragmatic inference should be to provide an account of local effects, and on this ground the standard Gricean model is rejected.

5. Alternate models and competing conceptions

In this section, we will discuss the views of theorists who, while fully embracing the idea that what is conveyed by an utterance includes both encoded content and inferentially derived content, differ from Grice in various ways. There are three principal parameters of difference:

(i) the rules or principles involved in inference

(ii) the nature of the input to inference and the interaction between encoded content and inferential content

(iii) the appropriate analysis of "what is said" or "literal meaning"

5.1. Explicature and Implicature in Relevance Theory

Relevance Theory (RT), first formulated in Sperber & Wilson (1986), is billed as a cognitive theory of communication. Conversational inference is a central feature of this theory. But Sperber & Wilson differ from the Gricean conception along all three of the parameters set out above.

On the RT view, inference is driven by a deep cognitive principle: the impulse to extract maximal useful information from a stimulus, balanced by the impulse to expend no more effort than seems justified by the anticipated benefits. (Here, we hear echoes of the Zipfian principles underlying Horn's Q- and R- principles.) RT uses the term *Relevance* (rather non-standardly) to characterize this balance between effort and pay-off: the more cognitive benefits – that is, useful information – a stimulus provides, the more Relevant it is; but this is offset by the processing effort required to derive the information. A stimulus has *optimal Relevance* just in case the cognitive benefits derived fully justify the cognitive effort required to derive them.

According to RT, linguistic communication is governed by the following principle:

Communicative Principle of Relevance

Every utterance conveys a presumption of its own optimal relevance.

The goal of interpretation is to identify a speaker meaning consistent with this presumption. This one over-arching principle is argued to do all the work which, in Grice's theory, is done by all the maxims combined.

So, one major difference between Grice and RT is the conception of what underlies conversational inference. A second is the specific principle which launches the process. The third major difference concerns the input to conversational inference and the interaction between encoded and inferential content.

On the RT conception, linguistic decoding produces an impoverished semantic object, a propositional radical or propositionally incomplete logical form. The first task of Relevance-driven inference is to fill out this skeletal object into a complete proposition. As noted above, Grice recognized that in order to identify the proposition expressed, an interpreter must fix the referents of referential expressions and resolve ambiguities, but said nothing as to the nature of the processes involved. RT argues, first, that what is required to construct a complete proposition from what is linguistically encoded goes well beyond reference fixing and disambiguation. Second, RT argues that there is no principled distinction to be made between processes involved in "filling out" a proposition and processes involved in deriving further inferences from the proposition so retrieved. All are driven by the same presumption of Relevance.

Inferentially derived content which goes towards "filling out" encoded content into a complete proposition is called *explicature*. Some aspects of content which Griceans treat as generalized conversational implicatures are analyzed in RT as explicatures, with a corresponding shift in the semantic analysis of the items in question. The interpretation of scalars is a central case.

Once all necessary explicatures have been derived, and the utterance has been associated with a proposition, Relevance-driven inference may continue. RT preserves the term *implicature* for these additional inferences, but makes a further subdivision within this class. In RT, the interpretation of an utterance is taken to proceed relative to a context, a context being simply some set of propositions retrievable by the addressee. The relevance of an utterance is partially determined by the number and utility of conclusions which can be deductively drawn from the context in conjunction with the proposition derived from the utterance. Part of the process of interpretation is the construction of an appropriate context. In RT, the term *implicature* is used both for propositions which are infered to be

part of the context for interpretation, and for deductive conclusions drawn on the basis of the constructed context. The former type of implicature is called an *implicated assumption* or *implicated premise*, the latter, an *implicated conclusion*. This distinction (not coincidentally) maps onto the distinction made above between background and foreground implicature.

In introducing the background/foreground distinction, we noted that the extension of the term *implicature* to background implicatures involved a departure from the Gricean conception of implicatures as part of speaker meaning. RT's implicated conclusions are also not necessarily part of Gricean speaker meaning. Suppose a speaker produces an utterance U, knowing that there are multiple contexts relative to which the interpreter might find U relevant, and knowing that U would give rise to different implicated conclusions in different contexts. Then there is no *particular* implicature which the speaker intends to be drawn, and perhaps not even a determinate set of candidates. For Grice, these inferences could not count as implicatures; for RT, they do. In fact, RT recognizes a continuum of implicatures, from those clearly meant by the speaker, to cases where the speaker expects *some* implicated conclusions to be drawn but does not know what they will be. However, inferences drawn by the interpreter but not driven by the requirements of Relevance lie beyond the limits of implicature.

5.2. Bach's Conversational Impliciture

Kent Bach's 1994 "Conversational Impliciture" does not present any revised theory of what underlies implicature derivation. The purpose of the paper is to demonstrate that "the distinction between what is said and what is implicated is not exhaustive" (p.124). Between the two lies a level which Bach calls *impliciture*.

Like the Relevance theorists, Bach takes it that the output of semantic decoding is typically propositionally incomplete. Propositional incompleteness drives a process which he calls *completion*: the filling in of "conceptual gaps" in a "semantically underdeterminate" sentence content. Completion is semantically or conceptually mandated. An interpreter, merely by virtue of their knowledge of the language, will be able to identify semantic incompleteness; and linguistic rules or forms will determine what sorts of completions are required.

But even after completion, we may not have arrived at the level of impliciture. In some cases, a further process called *expansion* takes place. This is best explained by illustration. Consider one of Bach's central examples: A mother says (unsympathetically) to her child who is complaining of a cut finger:

(37)   You're not going to die.

She is not telling her son that he is immortal – the literal meaning of the sentence – but merely that he will not die from the cut on his finger. A proper understanding of her utterance requires expanding the proposition literally expressed to one which is "conceptually more elaborate than the one that is strictly expressed." (p.135)   This process of expansion is driven by the interpreter's recognition that the speaker could not reasonably mean what she has literally said. So, like ordinary implicature, expansion is driven by the assumption of the general cooperativity of the speaker.

So, why are these not simply cases of implicature? Here is how Bach distinguishes the two notions:

> In implicature one says and communicates one thing and thereby communicates something else in addition. Impliciture, however, is a matter of saying something but communicating something else instead, something closely related to what is said (p.126)... Implicitures are, as the name suggests, implicit in what is said, whereas implicatures are implied by (the saying of) what is said. (p.140)

Recanati (1989, 2004 and elsewhere) offers a model of linguistic interpretation, similar in many respects to that of Bach. Like Bach, Recanati distinguishes conversational implicature from other types of pragmatic inference which operate locally and which do not utilize "full blown" Gricean reasoning. Like Bach, Recanati recognizes two distinct sub-types of lower-level pragmatic processes, which he calls *saturation* and *modulation*: these are roughly equivalent to Bach's *completion* and *expansion*. Only after these are completed can anythig akin to Gricean implicature generation begin.

5.3. Levinson: Generalized Conversational Implicature as Default Interpretation

Like Bach, Levinson is essentially a Gricean, but considers that Gricean theory requires modification to allow for the fact of "pragmatic intrusion:" inferential contributions to what is said. These contributions are what he identifies as Generalized Conversational Implicatures in Levinson 2000. The specific aspects of interpretation which he identifies as GCIs more or less overlap with those identified by Relevance theorists, Bach and Recanati as local pragmatic effects. Unlike them, Levinson proposes that GCIs can be derived on the basis of (elaborated versions of) the standard Gricean maxims of Quantity, Relation and Manner. In his formulation, these become the Q-, I-, and M-Maxims, respectively ("I" for informativity: Cf. Atlas & Levinson 1981).

Two aspects of the revised maxims are noteworthy. First, for each principle, Levinson gives both a Speaker's Maxim and a Recipient's Corollary. The latter articulates in some detail what sort of inferences are derivable given the interpretation rule. This detail is the second noteworthy aspect of the formulation. Unlike Horn's proposal, discussed above, which aims for maximum generality, Levinson incorporates into the speaker corollaries specific inference rules which produce the central GCI types with which he is concerned. Readers are referred to the original for the full principles.

As presented so far, Levinson's view looks like nothing more than a reformulation of Grice's. What distinguishes his position is his view as to how his principles apply. The central notion for Levinson is that of a *default*: the principles are claimed to be default "inferential heuristics" (p.35) which produce default interpretations. Levinson proposes that the GCI-generating principles apply automatically in the interpretation of any utterance, unless some contextual factor over-rides their application. The inference underlying GCIs is supposed to be "based *not* on direct computations about speaker-intentions but rather on general expectations about how language is normally used." (p.22). Application of default interpretation principles thus lies somewhere in between decoding of linguistically encoded content and the calculation of true implicatures.

The notion of default occurs in two ways in Levinson's view. First, there is the claim that the interpretative principles apply by default. Second, Levinson describes the output of these principles as "default interpretations." Yet as pointed out by Bezuidenhout & Morris (2004), the interpretations which arise are not defaults in the sense that their selection requires no choice among alternatives. Levinson himself, throughout the book, presents multiple options for GCIs which may be associated

with particular words or constructions. And it is clear, as emphasized by Bezuidenhout (2002), that the choice among these options often requires reference to wide context – and very likely, reference to considerations of speaker intention and standard Gricean principles.

As noted, a major motivation for Levinson's theory of GCIs is that these implicatures can attach to subordinate clauses, and contribute to what is said. Levinson claims that the nature of his rules is such that "the inference can be made without complete access to the logical form of the whole utterance…Procedurally, the expression *some of the boys G'd* can be immediately associated with the default assumption *some but not all of the boys G'd* even when some indeterminate aspect of the predicate G has not yet been resolved" (p.259). Applications of the I-principle, which for Levinson include enrichment of lexical content, might similarly be triggered as soon as a particular word is encountered e.g. the interpretation of "drink" as "alcoholic drink" in *I went out for a drink last night.* This local triggering of GCIs is crucial to Levinson's account of pragmatic intrusion. However, Bezuidenhout 2002 argues rather convincingly against this view of GCIs.

Levinson clearly holds that standard Gricean implicatures also arise, but gives no indication as to whether they are generated by the same principles which produce GCIs, applying in a global, non-default manner; or whether PCIs are assumed to be generated by the standard (but almost identical) Gricean principles.

56. Formal and Experimental Approaches

In this article, we have reviewed the foundations of the analysis of conversational implicature. To conclude, I will briefly introduce two formal approaches to the study of implicature: Optimality Theoretic Pragmatics, and Game Theoretic approaches.

The Optimality Theoretic approach to interpretation is presented as a new take on the Radical Pragmatics program, adopting the view that linguistic form underdetermines the propositional content of utterances. In this respect, OT pragmatics shares underlying assumptions with Relevance Theory, which is cited in some presentations. It is proposed that OT can provide a framework both for the processes involved in filling out a full propositional content (e.g. pronoun resolution, fixing domain restrictions for quantifiers) and in further elaboration of the content expressed, that is, the

generation of implicatures. OT pragmatics has so far focussed principally on generalized implicatures involving standard but defeasible interpretations of particular forms: for example, scalar implicatures,and the association of stereotypical situations with unmarked syntactic forms (e.g. the interpretation of *stop the car* vs. *cause the car to stop*) – Horn's "division of pragmatic labor." So far, highly particularized implicatures lie outside of the purview of the theory.

OT pragmatics take much of its inspiration from Horn's (1984) pragmatic taxonomy (see section 2.4 above), which posits the two conflicting principles, Q and R. Horn's proposal is interpreted within OT pragmatics as claiming that pragmatic interpretation is a matter of achieving an optimal balance in the pairing of forms and interpretations: a given form f should be assigned an interpretation m which enriches the linguistic content of f, but not so much that some alternate form f' would be a better candidate for the expression of m; while a desired meaning m should be expressed in a form f which is adequately economical, but not so much so that some other meaning m' would be a better candidate to be expressed by it. This balancing of competing requirements is captured by a modification of standard OT called Bidirectional OT (Blutner 2000).

Clearly what this conception requires is a theory of what renders one candidate form-interpretation pair better than another. Identifying a set of pragmatic constraints relative to which to formulate a metric for evaluating the "goodness" of such pairs is one of the central tasks of OT pragmatics. (For some contributions to this project, see Blutner & Zeevat (2004). The introduction to that volume gives a useful overview of the framework.) A second task of the theory is to fix the definition of optimization, i.e. what counts as an "adequately good" pairing. See again the aforementioned volume, and also Blutner (2000). For a book length presentation of the OT approach to interpretation, integrating syntax, semantics and pragmatics, see Blutner, de Hoop & Hendriks (2006). See also Article 103 *Optimality-theoretic pragmatics*.

Decision theory is a mathematical framework used to model rational decision making, in particular in situations of incomplete knowledge about the world. The decisions in question are typically choices among some fixed set of options. Game theory, an extension of decision theory utilizing the same set of formal tools, is used to model decisions in situations where two agents interact, and where each agent's choices impact on the decision of the other, a so-called *game*. Decision theory and, more extensively, game theory have been applied to various aspects of linguistic

analysis since Lewis's pioneering (1969) work. More recently, some theorists have begun to use game theory to model pragmatic inference, including the calculation of some types of implicature. (For a useful introduction to the basic mathematical machinery and its application to pragmatics, see chapter one of Benz, Jäger & van Rooij 2006; the rest of the book provides a good survey of central topics in the application of game theory to pragmatics. See also article 105 *Game Theory*.)

There are two (not entirely distinct) lines of work in this domain. One line of work attempts to formalize fundamental pragmatic concepts or principles using game theoretic tools. Particular attention has been given to the concept of Relevance (see e.g Marin 1999, van Rooij 2000). A second line of work attempts to replicate the effects of particular Gricean or neo-Gricean principles using Game Theory. Work in this line is generally presented as friendly to the Gricean framework, intended as a precise formulation of it rather than as an alternative.

Like OT, game theory is set up to model choices among a fixed set of alternative utterances/interpretations, and hence finds applications in attempts to model generalized implicatures involving choices among competing expressions. As with OT, particularized implicatures lie (at least so far) outside of the domain of the theory. Again, scalar implicature and Horn's "division of pragmatic labor" have been targeted for analysis (see van Rooij 2008 and article 105). The overlap between game theory and OT is not accidental. Dekker & van Rooij (2000) demonstrate that optimality theoretic models can be given a game theoretic interpretation, and that optimal interpretations represent Nash equilibria.

These are only samples of the formal approaches that have been utilized to develop robust accounts of conversational inference in general and conversational implicature in particular. A very different line of work, mainly pursued by computer scientists modelling discourse utilizes Planning Theory (for a starting point, see Grosz & Sidner 1986).

The developing area of experimental pragmatics offers another kind of enrichment of traditional Gricean pragmatics, applying the methods of experimental cognitive psychology to derive data for pragmatic theorizing. Some work in this area simply tries to apply proper methodology to elicit judgments from untutored informants, rather than using the standard informal (and, it is argued, often misleading) methods traditional in pragmatics: see for example Geurts & Pouscoulous 2009. Other work is aimed at testing the adequacy of particular pragmatic theories or claims by evaluating

their implications for processing. The approach is to formulate a specific claim about processing which is implied by the pragmatic theory under investigation, and then to test that claim empirically. This approach has been applied recently to the debate about the status of generalized conversational implicatures, with a variety of experimental techniques being applied to the question of whether or not these implicatures are "defaults" of some kind. Techniques range from the use of measures of reading or response time (e.g. Breheny, Katsos & Williams 2005), eye-tracking over visually presented examples (e.g. Bezuidenhout & Morris 2004) and even over real world tasks (e.g. Grodner & Sedivy 2004, Storto & Tanenhaus 2005). Noveck & Sperber 2004 offers a useful introduction to this area of research.

While an understanding of the conceptual foundations of the theory of implicature is crucial for meaningful work in this domain, it is only through the development of formal models that a substantive, predictive theory of conversational implicature can be provided.

6. References

Atlas, Jay 1977. Negation, Ambiguity and Presupposition. *Linguistics and Philosophy* 1: 321-336.

Atlas, Jay 1978. On presupposing. *Mind* 87: 396-411.

Atlas, Jay 2005. *Logic, Meaning and Conversation*. Oxford: Oxford University Press.

Atlas, Jay & Stephen Levinson 1981. *It*-clefts, informativeness and logical form: Radical Pragmatics (revised standard version). In P. Cole (ed.), *Radical Pragmatics*. New York: Academic Press.

Bach, Kent 1994. Conversational Impliciture. *Mind & Language* 9: l24-62.

Benz, Anton, Gerhard Jäger & Robert van Rooij 2006. *Game Theory and Pragmatics*. Basingstoke: Palgrave Macmillan.

Bezuidenhout, Anne 2002. Generalized Conversational Implicatures and Default Pragmatic Inferences. In Campbell, J., O'Rourke, M. & Shier, D. (eds.), *Meaning and Truth: Investigations in Philosophical Semantics*. New York: Seven Bridges Press, pp. 257-283.

Bezuidenhout, Anne & Robin Morris 2004. Implicature, Relevance and Default Inferences. In D. Sperber & I. Noveck (eds.), *Experimental Pragmatics*. Palgrave Press, 257-282.

Blutner, Reinhard 2000. Some aspects of optimality in natural language interpretation. *Journal of Semantics* 17: 189-216.

Blutner, R., Helen de Hoop & Petra Hendriks 2006. *Optimal Communication*. Stanford, CA: CSLI Publications.

Boër, S.E. & W.G. Lycan 1976. The myth of semantic presupposition. In A. Zwicky (ed.), *Papers in nonphonology* (*Working Papers in Linguistics* Vol.21). Columbus, Ohio: Department of Linguistics, Ohio State University.

Breheny, Richard, N. Katsos & J. Williams 2005. Are generalised scalar implicatures generated by default? An on-line investigation into the role of context in generating pragmatic inferences. *Cognition* 100:3, pp.434-463.

Chierchia, Gennaro 2004. Scalar implicatures, polarity phenomena and the syntax/pragmatics interface. In A. Belletti (ed.), *Structures and Beyond.* Oxford: OUP, 39-103.

Chierchia, Gennaro 2006. Broaden your views: implicatures of domain widening and the 'logicality' of language. *Linguistic Inquiry* 37, 535-590.

Cohen, Jonathan 1971. Some remarks on Grice's views about the logical particles of natural language. In Y. Bar-Hillel (ed.) *Pragmatics of Natural Languages*. Dordrecht: Reidel, 50-68.

Davis, Wayne 1998. *Implicature: Intention, Convention and Principle in the failure of Gricean Theory*. Cambridge: Cambridge University Press.

Dekker, Paul & Rob van Rooij 2000. Bi-Directional Optimality Theory: An application of Game Theory. *Journal of Semantics* 17: 217-242.

Fox, Danny 2007. Free choice and the theory of scalar implicatures. In U. Sauerland & P. Stateva (eds.) *Presupposition and Implicature in Compositional Semantics*. Basingstoke: Palgrave Macmillan, 71-120.

Franke, Michael, Tikitu de Jager & Robert van Rooij 2008. Relevance in Cooperation and Conflict. Ms, Institute for Logic, Language and Information, University of Amsterdam. To appear in Journal of Logic and Computation.

Gazdar, Gerald 1979. *Pragmatics: Implicature, Presupposition and Logical Form*. New York: Academic Press.

Geurts, Bart 2006. Implicature as a discourse phenomenon. In: Estela Puig-Waldmüller (ed.), Proceedings of Sinn und Bedeuting 11. Universitat Pompeu Fabra, Barcelona. Pp. 261-275.

Geurts, Bart 2009. Scalar implicature and Local Pragmatics. *Mind and Language* 24(1), 51-79.

Geurts, Bart & N. Pouscoulous 2009. Embedded Implicatures ?!? *Semantics and Pragmatics* Vol.2: 1-34. doi: 10.3765/sp.2.4

Green, Mitchell S. 1998: 'Direct Reference and Implicature.' Philosophical Studies 91, 61-90.

Green, Mitchell 1999. Illocutions, Implicata, and What a Conversation Requires. *Pragmatics and Cognition* 7: 65-91.

Grice, Paul 1975. Logic and Conversation. In D. Davidson & G. Harman (eds.), *The Logic of Grammar*, Encino, CA: Dickenson, 64-75. Reprinted in Grice 1989.

Grice, Paul 1978. Further notes on Logic and Conversation. In P. Cole (ed.), *Syntax and Semantics volume 9: Pragmatics*. New York: Academic Press, 183-97. Reprinted in Grice 1989.

Grice, Paul 1981. Presupposition and Conversational Implicature. In *Radical Pragmatics*. New York: Academic Press.

Grice, Paul 1989. *Studies in the Way of Words*. Cambridge MA: Harvard University Press.

Groenendijk, Jeroen & Martin Stokhof 1984. *Studies on the semantics of questions and the pragmatics of answers*. Ph.D. dissertation, University of Amsterdam.

Grodner, Daniel & J. Sedivy 2004. The effect of speaker-specific information on pragmatic inferences. In N. Pearlmutter & E. Gibson, (eds.), *The processing and acquisition of reference*. Cambridge, MA: MIT Press.

Grosz, Barbara J. & Candace Sidner 1986. Attention, intentions and the structure of discourse. *Computational Linguistics* 12(3), 175-204.

Hirschberg. Julia 1991. *A Theory of Scalar Implicature*. New York: Garland Publishing Company. (PhD. Dissertation, University of Pennsylvania 1985)

Horn, Laurence R. 1984. Toward a New Taxonomy for Pragmatic Inference: Q-Based and R-Based Implicature. In D. Schiffrin, (ed.), *Meaning, Form and Use in Context: Linguistic Applications*. Washington D.C.: Georgetown University Press.

Jäger, Gerhard 2008. Game theory in semantics and pragmatics. Ms., University of Bielefeld.

Katzir, Roni 2007. Structurally-Defined Alternatives. *Linguistics and Philosophy* 30(6), 669-690.

Kempson, Ruth 1975. *Presupposition and the Delimitation of Semantics*. Cambridge: Cambridge University Press.

Landman, Fred 1998. Plurals and maximalization. In S. Rothstein (ed.), *Events and Grammar*. Dordrecht: Kluwer, 237-271.

Landman, Fred 2000. *Events and Plurality*. Dordrecht: Kluwer.

Levinson, Stephen 2000. *Presumptive Meanings: The theory of Generalized Conversational Implicature*. Cambridge, MA: MIT Press.

Martinet, A. 1962. *A functional view of language*. Oxford: Clarendon Press.

Matsumoto, Yo 1995. The conversational condition on Horn scales. *Linguistics and Philosophy* 18(1), 21-60.

Merin, Arthur 1999. Information, relevance and social decision making: Some principles and results of decision-theoretic semantics. In L. Moss, J. Ginzburg & M. de Rijke (eds.), *Logic, Language and Information* Vol. 2. Stanford, CA: CSLI Publications.

Noveck, Ira & D. Sperber (eds) 2004. *Experimental Pragmatics*. Basingstoke: Palgrave Macmillan.

Recanati, Francois 1989. The pragmatics of what is said. *Mind and Language* 4, 295-329.

Recanati, Francois 2004. *Literal Meaning*. Cambridge: Cambridge University Press.

van Rooij, Robert & Katrina Schulz 2004. Exhaustive interpretation of complex sentences. *Journal of Logic, Language and Information* 13, 491-519.

van Rooij, Robert 2008. Games and Quantity Implicatures. *Journal of Economic Methodology* 15(3).

Sadock, Jerry M. 1976. On testing for conversational implicature. In In P. Cole (ed.), *Syntax and Semantics volume 9: Pragmatics*. New York: Academic Press, 183-97.

Simons, Mandy 2001. On the conversational basis of some presuppositions. In R. Hastings, B. Jackson & Z. Zvolensky (eds.), *Proceedings of Semantics and Linguistic Theory 11*. Ithaca, NY: CLC Publications.

Simons, Mandy 2004. Presupposition and Relevance. In Zoltán Szabó (ed.), *Semantics vs. Pragmatics*, Oxford University Press, 329-355.

Simons, Mandy to appear. "A Gricean view on Intrusive Implicature." In Petrus, Klaus (ed.), Meaning and Analysis: New Essays on Grice. (Palgrave Studies in Pragmatics, Language and Cognition Series.) Basingstoke: Palgrave, Macmillan.

Simons, Mandy 2008. Presupposition and Cooperation. Ms., Carnegie Mellon University.

Sperber, Dan & Deidre Wilson 1986. *Relevance: Communication and Cognition*. Cambridge, MA: Blackwell. Second Edition published 1995.

Stalnaker, Robert 1974. Pragmatic Presuppositions. In Milton K. Munitz & Peter K. Unger (eds.), *Semantics and Philosophy*. New York: New York University Press.

Storto, Gianluca & M.K. Tanenhaus 2005. Are scalar implicatures computed online? In Emar Maier, Corien Bary & Janneke Huitink (eds.), *Proceedings of Sinn und Bedeutung 9*. Nijmegen: Nijmegen Center for Semantics, pp.431-445.

Thomason, Richmond 1990. Accommodation, Meaning and Implicature: Interdisciplinary foundations for pragmatics. In Philip Cohen, Jerry Morgan & Martha E. Pollack (eds.), *Intentions in Communication*. Cambridge, MA: MIT Press.

Wilson, Deidre 1975. *Presupposition and Non-truth-conditional semantics*. London: Academic Press.

Zipf, G. K. 1949. *Human behavior and the principle of least effort*. Cambridge: Addison-Wesley.

Mandy Simons, Pittsburgh (USA)