Published in D. Sperber & I. Noveck (eds.), *Experimental Pragmatics*. Palgrave Press, 2004, 257-282.

Implicature, Relevance, and Default Pragmatic Inference

Anne L. Bezuidenhout and Robin K. Morris
University of South Carolina

Abstract

In his paper 'Logic and conversation' Grice (1989: 37) introduced a distinction between generalized and particularized conversational implicatures. His notion of a generalized conversational implicature (GCI) has been developed in two competing directions, by neo-Griceans such as Horn (1989) and Levinson (1983, 1987b, 1995, 2000) on the one hand, and relevance theorists such as Sperber & Wilson (1986) and Carston (1988, 1993, 1995, 1997, 1998a,b) on the other. Levinson defends the claim that GCIs are inferred on the basis of a set of default heuristics that are triggered by the presence of certain sorts of lexical items. These default inferences will be drawn unless something unusual in the context blocks them. Carston reconceives GCIs as contents that a speaker directly communicates, rather than as contents that are merely conversationally implicated. GCIs are treated as pragmatic developments of semantically underspecified logical forms. They are not the products of default inferences, since what is communicated depends heavily on the specific context, and not merely on the presence or absence of certain lexical items. We introduce two processing models, the Default Model and the Underspecified Model, that are inspired by these rival theoretical views. This paper describes an eye monitoring experiment that is intended to test the predictions of these two models. Our primary concern is to make a case for the claim that it is fruitful to apply an eye tracking methodology to an area of pragmatic research that has not previously been explored from a processing perspective.

Address for correspondence:
Anne Bezuidenhout
Department of Philosophy
University of South Carolina
Columbia, SC 29208
Tel (803) 777-3738
Fax (803) 777-9178
anne1@sc.edu

1. Introduction

Grice distinguished between generalized and particularized conversational implicatures. The latter he described as "cases in which an implicature is carried by saying that p on a particular occasion in virtue of special features of the context". The former he characterized as cases in which the "use of a certain form of words ... would normally (in the absence of special circumstances) carry such-and-such an implicature or type of implicature." (Grice 1989: 37). Grice did not develop the notion of a generalized conversational implicature (GCI) to any great extent. When he introduces the terminology in his paper 'Logic and conversation' he gives a few examples of the following sort:[1]

(1) A man came to my office yesterday afternoon.

(2) Max found a turtle in a garden.

(3) Robert broke a finger last night.

In the case of (1) the hearer would be surprised to discover that the man was the speaker's husband, for the use of the indefinite noun phrase 'a man' implicates that the speaker is not intimately related to the man. Similarly, in (2) we assume that neither the turtle nor the garden was Max's own, for if they were, the speaker would surely have used the expressions 'his turtle' and 'his garden'. On the other hand, the use of an indefinite noun phrase does not always implicate the lack of an intimate relation between the subject and the thing indicated by the noun phrase. In the case of (3) there is an implicature that it was Robert's own finger that Robert broke.[2]

Grice held that one mark of an implicature, whether generalized or particularized, is that it can be cancelled. A GCI will be explicitly cancelled when the speaker says something incompatible with the GCI. For example, the speaker of (1) could follow her utterance of (1) with the following assertion:

(1') I should confess that the man was my husband.

---

[1] At the end of 'Logic and conversation' Grice also suggests that the use of truth-functional connectives such as 'and' and 'or' give rise to GCIs. These suggestions of Grice's have been extensively and systematically explored by others, and will be discussed later under the heading of Q- and I-implicatures.

[2] Grice's discussion of examples like (3) is somewhat indeterminate. He could be interpreted as claiming that the utterance of (3) leads to the GCI that the broken finger was Robert's own. On the other hand, he could be interpreted as claiming that in the case of (3) the normal GCI associated with the indefinite (viz. the suggestion of a lack of intimate relation) is implicitly cancelled. Following Nicolle and Clark (1999) we will assume that there are at least two distinct types of GCI associated with the use of the indefinite, which can be labeled the alienable possession implicature and the inalienable possession implicature respectively. Fingers are inalienable possessions of people, so 'a finger' in (3) will lead to the GCI that the finger in question is Robert's own. Turtles on the other hand are alienable possessions of people, so 'a turtle' in (2) will lead to the GCI that the turtle in question was not Max's own.

A GCI will be implicitly cancelled when contextual information that is mutually manifest to both speaker and hearer is incompatible with the GCI. For example, if Robert is a known Mafia enforcer, then we may very well not derive the GCI that Robert broke his own finger. Or if we know that Max has a lot of pet turtles that went missing, and that he has been searching everywhere for them, we may not assume that the turtle he found was not his own.[3]

The notion of a GCI has been extensively explored by neo-Griceans, such as Atlas & Levinson (1981), Gazdar (1979), Hirschberg (1991), Horn (1984), and especially Levinson (1983, 1987b, 1995, 2000). A lot of attention has been paid to so-called scalar and clausal implicatures, which are sub-classes of what Levinson (2000) calls Q-implicatures. Another important class of implicatures is what Horn (1984) calls R-implicatures and Levinson (2000) calls I-implicatures. Levinson (2000) also identifies a third general class of GCIs that he calls M-implicatures. These three sorts of GCIs are derived from what Levinson calls the Q-, I-, and M-Principles respectively. Here we cannot give a detailed account of Levinson's treatment of these three classes of implicatures, or explain how Levinson thinks such implicatures are derivable from his three principles. We will rely on just a few examples to give some hint of the view under discussion. These examples do not cover the whole range of cases that have been classified as GCIs:

(4) Some books had color pictures.

GCI: Not all books had color pictures.

(5) We scored three goals.

GCI: We scored at most three goals.

(6) There might be a parrot in the box.

GCI: There does not have to be a parrot in the box.

(7) I believe that there is life elsewhere in the universe.

GCI: For all I know there may or may not be life elsewhere in the universe.

(8) Laurent broke his ankle and went to the hospital.

GCI: Laurent broke his ankle and then went to the hospital (This is a case of what Levinson (2000) calls conjunction buttressing).

---

[3] If one holds that the assumption that the man was not the speaker's husband is not an implicature of (1) but is part of what is said by the utterance of (1), then one may dispute that (1') is an explicit cancellation of a GCI. Rather, one will think of it as a retraction or reformulation of what was said.

(9) John caused the car to stop.

GCI: John brought the car to a halt in an unusual manner.

(4) – (6) are examples of scalar implicatures, (7) is a clausal implicature, (8) is an I-implicature and (9) is an M-implicature.[4]

Levinson thinks of all these kinds of implicatures as depending on metalinguistic knowledge. In the case of a scalar implicature, the hearer will not derive the GCI unless she is aware that the speaker has used an expression that is the weaker member of a so-called Horn or entailment scale. For instance, the expressions 'all' and 'some' form the entailment scale <all, some> because there is a one-way entailment relation between sentences (of an appropriate sort) that contain the quantifier 'all' and those same sentences with 'some' substituted for 'all'. More generally, an entailment scale is an ordered n-tuple of expression alternates $<\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n>$ such that if $\underline{S}$ is "an arbitrary simplex sentence-frame" and $\underline{x}_i > \underline{x}_j$, $\underline{S}(\underline{x}_i)$ unilaterally entails $\underline{S}(\underline{x}_j)$. (Levinson 2000: 79). When a speaker uses the weaker expression 'some' she implicates that she knows that the sentence with the stronger expression 'all' substituted for 'some' is false (or that she doesn't know whether the stronger statement is true). Clausal implicatures depend in a slightly different way on the existence of such entailment scales.

In the case of I-implicatures, the hearer must realize that the speaker has used an unmarked, minimal expression, which then licenses the hearer to use stereotypical information made available in the context to enrich the content of speaker's utterance. For instance, the connective 'and' is a minimal expression, and its use licenses the hearer to infer the enriched 'and then' interpretation. M-implicatures are in some sense the opposite of I-implicatures. They are licensed when the speaker uses a marked or prolix form of expression. For instance, the phrase 'caused the car to stop' is a marked way of speaking, and the hearer is thus licensed to infer that the speaker is suggesting that there was something unusual or non-stereotypical about the way the car was stopped. Clearly, in both the case of I-implicatures and the case of M-implicatures, speakers and hearers need to be aware that there are marked and unmarked ways of saying (roughly) the same thing, which is why Levinson thinks there is a metalinguistic element to such GCIs.

---

[4] Some people include more under the heading of GCIs than Levinson would be willing to include. Gibbs & Moise (1997), in some experimental work that is intended to test between a neo-Gricean view of GCIs and a rival view according to which GCIs are in fact a part of what is said, include cases of quantifier domain restriction under the heading of GCIs. E.g., the quantifier in 'Everyone is a vegetarian', might in context be understood to be restricted to everyone whom we have invited to dinner tonight. Gibbs & Moise also include what they call "time-distance" sentences under the heading of sentences that give rise to GCIs. These are cases like 'It will take us some time to get to the mountains' or 'The park is some distance away'. For instance, in context the former can convey the proposition that it will take longer than the hearer might have expected to get to the mountains. Levinson (2000) does not discuss examples like these, so it is unclear how he would deal with them. They could be assimilated to cases like conjunction buttressing, on the grounds that expressions like

Levinson proposes to treat GCIs as the result of "default pragmatic inferences which may be cancelled by specific assumptions, but otherwise go through." (Levinson 1987a: 723). He has been developing a set of heuristics or default inference rules that he says are used to generate GCIs. (Levinson 1987b, 1995, 2000). These default inferences yield interpretations that represent a level of meaning that he calls utterance-type meaning, which is intermediate between sentence-type meaning and speaker meaning.[5]

In this paper we will use Levinson's account of GCIs as the basis for our own speculations about how sentences that give rise to GCIs are processed. We also contrast the processing view derived from Levinson with an alternative processing account derived from the work of Sperber & Wilson (1986), Carston (1988, 1993, 1995, 1997, 1998a,b), and Recanati (1991, 1993). This alternative account sees GCIs as interpretations that are arrived at by the pragmatic development of semantically underspecified logical forms. This development occurs as the result of the operation of local pragmatic processes of enrichment and loosening. A hearer trying to understand a speaker's utterance will use the information that is semantically encoded in the words the speaker uses along with information that is mutually manifest to speaker and hearer in the conversational context. The lexical concepts the hearer accesses will be pragmatically enriched or loosened to yield ad hoc concepts. These ad hoc concepts become constituents of a representation the hearer is building of the proposition expressed by the speaker's utterance. In other words, these enriched meanings are attributed to the speaker as part of what the speaker said, not as something merely implicated. In this process the Gricean notion of what is strictly and literally said plays no role. Enriched interpretations are directly constructed via the local processes just described.

Clearly what we have said so far cannot do justice to the rich and detailed accounts that have been offered by these rival theorists on the topic of GCIs. We have limited our description to those aspects of the theory that are relevant to two very simple pragmatic processing models, the Default Model (DM) and the Underspecification Model (UM), which are loosely inspired by the neo-Gricean and Relevance

---

'everyone', 'some time', and 'some distance' are unmarked, minimal expressions that invite contextual enrichment. In other words, they would be cases falling under Levinson's I-Principle.

[5] Levinson's account is similar to the one that has been developed by Kent Bach (1994a,b, 1999), although the terminology Bach uses to describe his views is very different from Levinson's. To articulate the similarities and differences between these views would require a separate paper, but they seem to be in agreement about the default status of GCIs. What Levinson calls GCIs Bach calls implicitures, but just as Levinson thinks that GCIs are the result of default inferences, Bach thinks implicitures are derived on the basis of default inferences, which in turn are the result of a process of standardization.

Theory views respectively.[6] Both these models are intrinsically interesting, and that both have some a priori appeal. But we consider them to be incompatible with one another. They cannot both be true. In section 3 we report data from an eye monitoring experiment that provides some initial data regarding on-line processing of sentences that give rise to GCIs and discuss how these data may inform further development of processing models derived from these two theoretical perspectives. Before proceeding to the experiments we flesh out the models in more detail.

2. Two competing pragmatic processing models

As we understand them, DM and UM are rival models of the pragmatic processes that are involved in understanding utterances of the sort that Grice claimed give rise to GCIs. One simple way to see how DM and UM differ is to look at a couple of examples and see the different accounts these models give of the process that leads to the recovery of GCIs. We will look first at the case of scalar expressions, which fall under Levinson's Q-Principle.

According to DM, expressions that give rise to scalar implicatures, like 'some', 'three', 'possibly' and so on, belong to entailment scales. An entailment scale is a set of lexicalized expressions that belong to the same semantic field and that are ordered according to 'strength'. For example, <all, many, some> is an entailment scale, as are < …, four, three, two, one> and <necessarily, possibly>. According to DM, the use of a sentence containing a weak expression from such an entailment scale gives rise to a default GCI. It implicates the denial of that same sentence with any stronger expression from the same scale substituted for the weaker one. Thus 'Some dogs have spots' scalar implicates 'Not all dogs have spots' (as well as 'Not many dogs have spots').

Let us look in more detail at the case of cardinal sentences such as:

(10) Jane has three children.

a. Jane has at least three children.

b. GCI: Jane has no more than (i.e., at most) three children.

c. Jane has exactly three children.

According to DM 'three' means 'at least three'. Hence what is said by (10) can be represented as (10a). The Q-Principle will be triggered by the cardinal expression 'three' in (10), which belongs to the

[6] There is some theoretical value to testing extreme versions of theoretical hypotheses. For one thing, these extreme versions make stark predictions, which may not follow from the more hedged theories that are their inspiration.

entailment scale < … four, three, two, one>, thereby yielding the default GCI (10b).[7] What is said together with the GCI thus entails (10c). This process of retrieving the default GCI and combining it with what is said would be DM's explanation for why in many contexts the expression 'three children' in (10) is understood to mean exactly three children.

UM on the other hand claims that expressions such as 'some' and 'three' are semantically underspecified. They must be specified in context and how they are specified depends on the operation of a local pragmatic process of enrichment. In the case of (10), a pragmatic process of enrichment takes the semantically underspecified lexical concept three, and yields a contextually appropriate enrichment. Depending on the assumptions accessible in the context, the proposition expressed by (10) could be either (10a) or (10b) or (10c). In particular, to understand the speaker to have been communicating (10c), the hearer need not go through a process whereby (10a) and (10b) are retrieved as well.

The two models DM and UM also treat cases falling under Levinson's I-Principle differently. Consider the case of what Levinson calls conjunction buttressing:

(11) Laurent broke his ankle and went to the hospital.

a. (Laurent broke his ankle) $\wedge$ (Laurent went to hospital).

b. GCI: Laurent broke his ankle and then he went to the hospital.

According to DM, 'and' has the same meaning as the conjunction symbol '$\wedge$' in truth-functional logic. A consequence is that as regards what is said by a conjunction, it can be interpreted simply as a list, the order of the conjuncts being irrelevant to its overall meaning. Since 'and' is an unmarked, minimal expression, DM claims that it triggers the I-Principle and gives rise to the default GCI in (11b). In other words, DM holds that a temporal interpretation of the conjunction 'and' is the default interpretation.

On the other hand, UM claims that 'and' is semantically underdetermined and it can be specified in many different ways, depending on the assumptions that are operative in the discourse context. See Carston (1988, 1993). There will be cases in which it is appropriate to interpret 'P and Q' to mean P and then Q. But in other cases it may be appropriate to interpret it just as a list, or in some alternative enriched way as either:

P and as a result Q (e.g., 'She kicked him and he cried.')

P and for that reason Q (e.g., 'She cheated him and he sued her.')

---

[7] Strictly speaking according to the Q-Principle, the utterance of (10) would generate a whole series of GCIs 'Jane does not have four children', 'Jane does not have five children', and so on, which can be subsumed under the summary GCI represented by (10b).

P and simultaneously, Q (e.g., 'She took a shower and practiced her singing.')

P and in the course of doing P, Q (e.g., 'She talked to him and found she liked him.')

etc.

DM of course can admit that all these alternatives are possible interpretations, in the sense that the default and then interpretation can be cancelled, either implicitly or explicitly in favor either of the "logical" list meaning or one of these other enriched interpretations. According to DM, these alternative enriched understandings will not of course be GCIs, but particularized conversational implicatures (PCIs).[8]

One obvious difference between these models is that DM posits an additional step in the recovery of what Griceans think of as the GCI. First the hearer must recover what is said by the utterance, and this will trigger a further inference to the GCI. For instance, on hearing (10) a hearer will first recover the meaning Jane has at least three children and only then will he recover the GCI Jane has at most three children. UM on the other hand suggests that if the interpretation Jane has at most three children is constructed, it will be directly constructed by adjusting an underspecified lexical concept, and the hearer will not first need to construct the interpretation Jane has at least three children. In such a situation, the only candidate for what is said is Jane has at most three children, given that DM's candidate for what is said is not recovered. Similar remarks could be made with respect to sentences such as (11).

There has been some experimental work addressed to the question of whether DM's candidate for what is said must be constructed in the course of recovering the GCI. See Gibbs & Moise (1997); Nicolle & Clark (1999); Bezuidenhout & Cutting (2001). In the experimental literature, DM's candidate for what is said is, somewhat misleadingly, called the minimal meaning and the GCI is called the enriched meaning. The evidence overall seems to count against DM on this issue, since it has been shown that hearers are relatively insensitive to minimal meanings as compared with enriched meanings. This has been shown in off-line judgment tasks, where people, when asked to choose the meaning that they think best represents what is said, favor the enriched over the minimal meaning. It has also been shown in on-line self-paced reading tasks. These have shown that people are relatively slow to read sentences in contexts requiring only the minimal interpretation as compared to the time to read those same sentences when they are presented in contexts requiring that they be given enriched interpretations. If the minimal

[8] Another possible understanding of DM is that it is claiming that all these interpretations are default GCIs and hence all are accessed whenever the minimal expression 'and' is encountered. Context then selects the appropriate interpretation. We are assuming that this is not the version of DM that is most plausible, as we explain below.

meaning must always be accessed first, reading times in contexts favoring enriched readings should not be faster than reading times in contexts favoring minimal readings.[9]

In this paper we wish to focus on another aspect of DM; that is DM's claim that GCIs have the status of default inferences that are automatically triggered unless something in the context blocks them. We report on an eye monitoring experiment that is designed to explore the role of defaults in understanding sentences of the sort that Grice claims give rise to GCIs.

However, before turning to a description of our experiment, we believe that some clarification of terms is in order. In particular, we wish to explain how we understand the notion of default inference that is invoked by DM. Also, we try to clarify the notion of semantic underspecification that is invoked by UM.

2.1 Default inferences

What is generally meant by calling something a default is that it settles some issue without the need to make a choice among alternatives. For instance, consider what is meant by calling something a default assumption. Suppose we are in a restaurant and have been shown to a table and given some menus. We assume that a waiter or waitress will come to the table to take our order. This is a default assumption, and we don't have to think about alternative scenarios, unless something untoward happens. Something similar holds in the case of default inferences. For example, normally when something looks red to me I will infer that it is red. This is a default inference. It is one that will be drawn, all things being equal. If all things are not equal, say because I am alerted to the fact that there is something unusual about my perceptual circumstances, then I will consider alternative conclusions. E.g. perhaps the object is actually white but I am seeing it through a red filter or perhaps the object is actually green and I am seeing it through color inverting spectacles, etc.

In several places Levinson suggests that a given sentence-type will be associated with several different GCIs. Take a sentence such as 'John's book is on the table'. Levinson (2000: 37, 207) grants

[9] Kent Bach (1995, 1998) has argued that by a process of standardization, the enriched meanings (i.e., the GCIs) have become entrenched. Thus when a hearer processes a sentence such as (10) the minimal meaning (what is strictly and literally said) can be bypassed. In fact, Bach thinks this is exactly what it means to say GCIs are defaults. We attend to what it literally said only in unusual circumstances, when something in the context signals that the default should not be drawn. But in normal circumstances the complete inferential process that first derives what is said and then derives the GCI will be short-circuited. This standardization explanation works well in some cases, when there is a formulaic GCI associated with a certain lexical item. But where two or more enrichments are possible, which vary with the context, then the standardization explanation is more problematic. There will be no standard GCI associated with a sentence type. The standardization view could be supplemented with some sort of frequency based account. The standard GCI could then be regarded as the one most frequently associated with a certain sentence type. We have more to say below about the possible role of frequency information in processing sentences of the sort that we are focusing on.

that by his I-Principle each of the enrichments 'book John owns/bought/borrowed/read/wrote', as well as many more, are possible for 'John's book'. Similarly, Levinson (2000: 38, 117) allows that temporal, causal and teleological interpretations are all possible under the I-principle for a sentence such as 'Jane turned the key and the engine started.' And multiple possible GCIs seem possible also under his Q-Principle. For instance, a scalar sentence such as 'I ate some of the cookies' ought to generate the GCIs 'I did not eat all/most/many of the cookies'. See Levinson (2000: 77).

However, if we allow multiple GCIs for a given sentence-type, and allow that they are all inferred whenever a sentence of that type is used, this will defeat the whole purpose of having a default in the first place. For the hearer will now be forced to decide between the multiple interpretations that have been inferred, since in many cases these interpretations will be incompatible with one another (E.g., John's book can't be both bought and borrowed). But, as already noted, the point of having a default is precisely to avoid the need for making choices unless forced to do so by unusual circumstances.

Levinson could retreat to the claim that although there are potentially many GCIs only one will actually be inferred and which one is actually inferred will be determined by context. However, if he adopts this position then it is no longer clear that we can talk about a system of default inferences. Such a view of defaults would be compatible also with UM, since all parties can agree that a sentence such as 'John's book is on the table' has many possible interpretations, including 'The book John owns/wrote/borrowed/bought/sold is on the table'. Which one of these interpretations will be selected depends on the context.

Thus it seems that to defend a view that can truly be said to be a default theory, we must select one from among all these alternatives as the single default. There are places where Levinson does seem to opt for a single default. For instance, in the case of conjunctions using 'and' Levinson (2000: 123) suggests that the default is the temporal understanding 'and then'. The problem with this is that it is unclear that this sort lack of flexibility in pragmatic processing is ultimately a good thing. It may speed up processing in some cases, but it could substantially hinder processing in other cases. For instance, if the default for 'John's book' is the ownership interpretation, then if the speaker means to communicate that the book is one that John borrowed, this will cause some difficulties in interpretation, as the default will first have to be cancelled/ overridden/ suppressed. This will require processing resources, thus increasing the hearer/reader's processing load. Hence the system of defaults could be said to be inefficient from a processing perspective.

Of course Levinson will stress that a default inference is one that will be drawn only if things are otherwise normal. If there is something unusual about the situation, the default will not be drawn (just as when I am alerted to something unusual in my perceptual circumstances I do not infer from something's

looking red that it is red). So Levinson might argue that if the context makes it clear that what we are talking about is the book John borrowed, then the default inference that the book is one that John owns will not be drawn. Hence there is no processing cost associated with setting the default to the ownership interpretation. This is so, provided that the context that supports the borrowing interpretation precedes the use of the phrase 'John's book'. If it does not, then the default inference will be drawn and there will then be a need to cancel/suppress/override the default.[10]

It has been suggested to us that there is no reason to deny that there can be multiple default GCIs associated with a sentence type. Just as an ambiguous expression is associated with many different meanings, all of which may be initially activated when the expression is used, so too a sentence type can be associated with multiple GCIs, all of which are inferred when the sentence is used. However, we do not think that ambiguous expressions provide a good analogy for thinking about sentences that give rise to GCIs. GCIs, even on the default model, are not ready-made chunks of meaning that are accessed from the lexicon. They must be computed, even if the computation is fast and automatic. For instance, it is not part of the lexically encoded meaning of 'John ate some of the cookies' that John did not eat all of the cookies. This is something that must be inferred, even if, as DM claims, this inference is a default one. Thus we continue to think that it is problematic to talk of multiple defaults. Moreover, the fact that ambiguity is not a good model for GCIs also means that lexical ambiguity resolution is not a good model for the process whereby a particular GCI is derived.

2.2 Semantic underspecification

The notion of semantic underspecification has been widely discussed in recent times, particularly by cognitive linguists such as Fauconnier (1985, 1997), Pustejovsky (1995, 1998), Van Deemter (1996), Coulson (2001), Tyler & Evans (2001) and others. It is also a crucial component of Sperber & Wilson's (1986) Relevance Theory. These theorists all stress that it is necessary to distinguish between the meaning that is encoded by a lexical expression and the interpretation that the item receives in an utterance context. Such interpretation frequently goes beyond the encoded meaning. It must be constructed on-line, on the basis of the encoded meaning of the expression together with non-linguistic contextually available information. This sort of meaning construction is a process of conceptual integration, which combines linguistic and non-linguistic concepts according to general cognitive principles. A semantically underspecified lexical expression is thus one whose encoded meaning does not fully specify its contextual meanings.

---

[10] Besides, someone who believes in a system of defaults shouldn't stress too much those contexts that block the generation of defaults. For if situations in which default inferences are blocked occur very frequently, the rationale for the idea that we need defaults in the first place is weakened.

In order to fully flesh out a processing model that relies on the notion of underspecification, one would have to know a lot more about the cognitive principles that are involved in sense construction. Moreover, one needs to have a clearer idea of what a semantically underspecified meaning is. One suggestion is that an underspecified meaning is the meaning that is common to all its possible specifications. This idea doesn't seem very plausible. For example, consider a polysemous noun such as 'newspaper'. It can refer to either a publisher ('The newspaper fired its editors') or a publication type ('The newspaper today has an obituary for Nozick') or a publication token ('The newspaper is on the kitchen table'). One might want to argue that the term itself is semantically underspecified, and that the various meanings it has are constructed from the encoded meaning on the basis of contextual information. However, it is not plausible to say that the encoded meaning in this case corresponds to the meaning that all these uses have in common. The different uses refer to things that belong to different ontological categories, and thus there may be nothing they share in common (except something trivial, such as being a thing, or something unhelpful, such as being able to be referred to by the term 'newspaper').

A more promising idea is to select one of the meanings as primary, and hence to regard its referent as the primary referent of the term. Other meanings will be secondary, and their referents will be things that stand in certain relations to the primary referent. Context will be needed to figure out what the relevant relation is in a particular case. For example, suppose one holds that 'newspaper' refers primarily to a publication token, then the underspecified meaning of 'newspaper' would be thing(s) that stands in relation R to this token publication. In context, the relation R could be specified as the identity relation, in which case the term refers to a publication token. Alternatively, it could be specified as the relation of publishing, in which case the term refers to the publisher of some publication token. This account is very similar to the one proposed by Recanati (1995) for possessive phrases such as 'John's book'. Recanati proposes to analyze this as making reference to a relation R holding between John and the book that has to be specified in context. If the relation is specified as one of ownership, then 'John's book' refers to the book John owns. If the relation is specified as one of authoring, then the phrase 'John's book' refers to the book John wrote, and so on.

In the course of their own work on the preposition ‘over’, Tyler & Evans (2001) lay out some helpful methodological principles to be followed in identifying the primary sense associated with such a polysemous term. They also suggest strategies to be followed in determining whether a sense is a distinct sense instantiated in semantic memory or is instead to be accounted for by a cognitive process of conceptual integration (i.e., is instead a meaning that must be constructed on-line). These methodological principles could be profitably applied in the current context, in order to give an account of the primary senses associated with terms such as ‘some’ and ‘and’, and to distinguish such primary meanings from

senses that must be constructed on-line by a process of conceptual integration of lexical and non-lexical concepts.

3. An eye movement monitoring study

In this section we report on an eye monitoring experiment that we hope can clarify the role of defaults in pragmatic processing. The experiment is focused on the processing of sentences giving rise to scalar implicatures of one particular sort. (In other work of ours not reported here we have focused on the processing of sentences that give rise to I-implicatures). Our experiment is focused on the processing of sentences of the form 'Some S are P', which, according to DM, give rise to the default GCI 'Not all S are P'. Such a GCI is an example of a scalar implicature, and according to DM is derived by means of the Q-principle.

Noveck (2001) has a very interesting study of scalar implicatures of this sort that some might read as lending support to DM. Noveck's study was aimed at uncovering a developmental trend in children's acquisition of certain modal and quantificational constructions. But the results he got for his adult control subjects are what are of interest to us here. Consider the following example:

(12) Some elephants have trunks.

a. Some but not all elephants have trunks.

b. Some and possibly all elephants have trunks.

c. Not all elephants have trunks. (GCI)

59% of adults in Noveck's study judged that sentences such as (12) are false. This means these adults must have interpreted (12) as (12a). DM has a ready explanation for this. Adults on hearing (12) automatically derived (12c), which is the scalar implicature triggered by the use of 'some' in (12). This GCI is combined with (12b), which is what is strictly and literally said by (12) according to DM, and thus people construe the speaker of (12) to have asserted (12a). This explanation requires DM to say that when people are asked to judge the truth or falsity of a statement, they actually end up judging the truth or falsity of something that is an amalgam of what is said and what is implicated.

Noveck himself does not construe the results of his experiments as lending support to DM. He says: "Our evidence is compatible with either the neo-Gricean or Relevance account of scalar implicatures." (2001:186). We think that Noveck is right to be cautious, since defenders of UM would presumably want to give their own explanation for Noveck's findings. According to UM when a speaker utters (12), the hearer will directly generate either the (12a) or (12b) reading from an underspecified form, and which one is generated will depend on the context. In particular, to generate the (12a) reading does not require first

generating the (12b) and (12c) readings, as DM claims. So if adults interpreted (12) as (12a) in Noveck's experiment, there must have been something in the context that made that the most relevant reading.

The main point we wish to make here is that post hoc processing explanations can be given for Noveck's data both from the point of view of the DM model and of the UM model, even if this is perhaps more difficult from the UM perspective. This suggests that we need processing data to differentiate between these two models. Eye movement monitoring during reading is one way to collect data of the required sort. The eye movement data provide an on-line record of processing as it unfolds over time without having to rely on secondary task responses gathered after the fact.

There is a large literature validating the use of fixation time measures to assess higher order cognitive processing (see, Rayner & Morris, 1990 for review). This literature has established a tight (albeit not perfect) link between where a reader is looking and what the reader is processing. For example, lexical factors such as word frequency and lexical ambiguity directly influence the initial processing time that readers spend on a word. Recent work by Poynor & Morris (under review) demonstrates that the gaze duration measures gleaned from the eye movement record are sensitive to the process of generating inferences. In situations in which readers have committed to a particular syntactic or semantic analysis and then find that they must abandon this analysis in order to successfully comprehend the text, readers may spend additional processing time on the new information and/or the reanalysis process may be reflected in second pass reading time on critical words or phrases (See also Folk & Morris, 1995).

3.1 Scalars, cancellation and predictions of the two models

If we are to be able to detect the presence of a default inference, we need a situation in which a default inference will be generated, but in which the reader subsequently gets information that is inconsistent with the default, thereby suggesting the need for a retraction of the meaning assignment made on the basis of the default inference. If a default is generated and then retracted, there should be evidence of processing difficulties in the eye movement record. In our experiment, readers saw pairs of sentences. The first sentence of each pair was of the form 'Some N were/had P'. It was followed by a sentence that explicitly cancelled the GCI 'Not all N were/had P'.[11] The second sentence was of the form 'In fact all of them were/did, …'. The word 'all' in the cancellation sentence is the first information that readers encounter that suggests that some but not all may not be the appropriate interpretation for the initial determiner phrase. The immediately following phrase 'them were/did' definitively rules out the some but not all interpretation. We then created two further versions of each item, by replacing the word 'some' in

[11] For the distinction between explicit and implicit cancellation of a GCI, see section 1 above.

the first sentence by either 'many' or 'the'. An example of the three versions of a typical item is given below:

(13)a. Some books had color pictures. In fact all of them did, which is why the teachers liked them.

b. Many books had color pictures. In fact all of them did, which is why the teachers liked them.

c. The books had color pictures. In fact all of them did, which is why the teachers liked them.

According to DM, the GCI 'Not all N' should be triggered automatically both when the reader sees 'Some N' and when s/he sees 'Many N' but not when s/he sees 'The N'. In general, a default is an alternative that we assume to be true unless we are told otherwise. Defaults are useful in that they allow us to proceed with processing in the face of incomplete information, and to avoid costly processing of multiple alternatives. However, this also implies a cost to the reader when those assumptions turn out to be wrong. In those conditions in which the GCI is derived, two things will occur when the reader gets to information that explicitly cancels the GCI. First, the reader must cancel the default GCI and second, they must reanalyze their discourse representation in order to retrieve or construct the context appropriate interpretation. It does not make sense that readers would engage in this costly cancellation and reanalysis process until and unless they are presented with compelling evidence of the need to do so. In our materials the information that explicitly cancels the GCI occurs when the reader encounters the phrase 'them were/did'. It is only when the reader processes the anaphoric pronoun, which refers back to the things N referred to in the initial sentence, that s/he can know that there is an explicit contradiction between the GCI ‘not all N’ and the 'all N' claim of the cancellation sentence. When readers encounter the word 'all', it is possible that the 'all' here is a quantifier applying to something N*, different from the N referred to in the initial sentence. This applies even if readers get a preview of the word 'of' when reading 'all'. For example, the sentence pair could have been:

(13a*) Some books had color pictures. In fact all of the pictures were highly colored, which is why the children liked them.[12]

Hence DM predicts that there will be increased processing time in the ‘them were/did’ region in the ‘Some N’ and ‘Many N’ conditions compared to the ‘The N’ condition. This increased processing time should show up in first pass time in this region. DM also predicts that there should be some indication

[12] We did construct 30 of our filler items to begin with some/many/the N sentences and then to continue in the divergent way that (13a*) does, instead of ending with a cancellation of a GCI. However, there were too few of these divergent filler items with the word 'all' in the continuation sentence to allow for a comparison with our 'Some N' condition. This is a comparison we intend to explicitly test in a follow-up study. We note also that the fact that we had had 30 filler items that started in the same way as our cancellation sentences but that did not end in a cancellation of the GCI 'not all' makes an Early Cancellation strategy very risky. If readers anticipated a cancellation every time they saw the words 'Actually', 'In fact', and so on, half the time they would be mistaken.

that readers are engaging in reanalysis. This could come in a number of different possible forms. The increased processing time on 'them were/did' could spillover onto the following region of the sentence (which we call the end-of-sentence region, viz. the region immediately following the phrase 'them were/did' to the end of the sentence). Or the reader might return to the initial determiner phrase (which we call the 'Det N' region) in the 'Some N' and 'Many N' conditions. This could be observed in the number of regressions or in rereading time on that initial region.

In contrast to DM, UM claims that readers in the 'Some N' condition do not fully commit to the some but not all reading right away. Rather, they engage in an incremental process utilizing all available information at any given moment in time. Under this view readers rely on probabilistic information to develop their interpretation over time. Thus, according to UM readers may (or may not, depending on your point of view) be biased toward the some but not all interpretation when they encounter the word 'some' but the item remains underspecified (in either case) until more information accrues. The word 'all' provides information that is biased toward the some and possibly all interpretation of 'some'. This predicts increased processing time on the word 'all' to reflect the fact that the reader has registered information potentially relevant to the specification of an underspecified item.[13] When readers reach 'them were/did' that information is consistent with their current interpretation. Thus there should be no increase in processing time in this region. There is no need for a reanalysis, since the incorrect interpretation never was assigned to 'some' and hence under this view there is no prediction of increased rereading time or regressions to the initial 'Det N' region. This account makes no firm predictions about differences in behavior between the 'Some N' and 'Many N' conditions. What probabilistic information is deemed relevant to the specification process and at what points it is deemed relevant may differ for 'some' and 'many'.

These predictions are summarized in Table 1 below:

[13] We are presupposing here that the specification process uses resources to integrate the new information, and hence that it has some processing cost.

Table 1: Summary of the predictions made by the various versions of DM and UM

Behavior in Some N condition and comparisons with and between controls	DM's predictions	UM's predictions
Increased time on all?	No	Yes
Increased time on them were/did?	Yes	No
Regressions/rereading of Some N?	Yes	No
Some should behave like Many?	Yes	??
Some should behave like The?	No	No
Many should behave like The?	No	??

3.2 Participants

24 participants from the University of South Carolina community were recruited for this experiment. They either received one experimental credit in a psychology course or were paid $5 an hour for their time. All participants had normal, uncorrected vision and were native speakers of English.

3.3 Materials and Design

We created a series of 30 items, each consisting of a pair of sentences. The first sentence of each pair was of the form 'Some N were/had P'. It was followed by a sentence that explicitly cancelled the GCI 'Not all N are P'. Two additional versions of each item were then created by replacing the word 'some' in the first sentence with either 'many' or 'the'. An example of the three versions of a typical item is given in section 3.1 above. The preface to the cancellation sentence was not always the same for all items. In constructing our materials we used a variety of phrases such as 'in fact', 'actually', 'as a matter of fact', 'in all truth', 'truth be told', 'of course' and 'in reality'. All participants saw 10 items from each of the 3 conditions (Some/Many/The). No one saw the same item in more than one condition. In addition, each person saw 63 filler items. 33 of these fillers came from another, unrelated experiment (whose materials also consisted of sentence pairs). The other 30 fillers were created to begin in the same way as the experimental items. However, what followed the preface to the second sentence in these items was not a cancellation of the GCI associated with the first sentence, but was either an elaboration of the first sentence or the introduction of some new topic. (E.g., 'Some/Many/The pictures were fuzzy. Actually their resolution was so bad that they were impossible to make out.'). Materials were presented in random order, a different order for each participant.

3.4 Procedure

Items were displayed on a color monitor 35 inches from participants' eyes. Text was centered on the screen, with 3 characters subtending 1° of visual angle. We monitored the movements of participants' right eyes during reading, using a Fourward Technologies Dual Purkinje Generation 5 eye-tracker. The software sampled the tracker's output every millisecond to determine fixation location within a single character position. A bite bar was prepared for each participant to minimize head movement. The experimenter calibrated the tracker for each participant and performed a check before each item. Participants were instructed to read for comprehension. A yes/no question (without feedback) followed 20% of the items. All participants performed at 80% or better on comprehension questions.

3.4 Data Analysis

We measured the time spent reading the phrase 'some/many/the N' (the Det N region) in the initial sentence. We also measured the words 'all' and 'them were/did' in the cancellation sentence, as well as the end-of-sentence region in the cancellation sentence. For single words we measured gaze duration, and second pass time. For multi-word regions we measured first pass time and second pass time. We also measured the number of regressions into (i.e., the number of times participants looked back at) the phrase 'some/many/the N', as well as regressions into the words 'all' and 'them were/did'. First pass time or gaze duration is all the time spent in a region before exiting to either the left or the right of that region.[14] Second pass time includes all the time spent rereading in a region, excluding gaze duration or first pass time. The first two measures give some indication of early processing, whereas second pass time and number of regressions into a region allow one to make inferences about the reanalysis and text integration processes. See Rayner, Sereno, Morris, Schmauder & Clifton (1989). We report ANOVAs treating participants (F_1) and items (F_2) as random effects.

3.6 Results

Table 2 below shows mean reading times (in milliseconds) for two of the regions of interest in the "cancellation" sentences in each of the three conditions (Some/Many/The N).

[14] 'Gaze duration' is the term used to refer to initial time spent in a one-word region whereas 'first pass time' is used to refer to initial time spent in a multi-word region.

Table 2: Mean initial reading times (in msecs) for 2 critical regions in "cancellation" sentences

	Gaze Duration on 'all'	First Pass Time on 'them were/did'
Some N	**275**	301
Many N	260	308
The N	256	**329**

The region 'all': Readers spent more initial processing time (as measured by gaze duration) on the word 'all' following 'Some N' than following 'The N' (F_1 (1,23) =10.59, MSE = 423, $p < .01$; F_2 (1,60) = 7.36, MSE = 1001, $p < .01$). This is consistent with the predictions of UM. DM predicts no increased processing time in the 'Some N' condition until the 'them were/did' region. There was no difference in gaze duration on 'all' between the 'Many N' and 'The N' conditions (Both F_1 and $F_2 < 1$). The lack of reading time difference between the 'Many N' and 'The N' conditions is also inconsistent with DM as DM predicted that the 'Many N' and 'Some N' conditions would produce similar effects relative to the 'The N' condition. UM made no prediction regarding this comparison.

The region 'them were/did': Initial processing time (as measured by first pass time) on the words 'them were/did' in the 'Some N' condition is faster than in the 'The N' condition, although this difference is reliable only in the subjects analysis. (F_1 (1,23) =11.44, MSE = 828, $p < .01$; F_2 (1,60) = 3.16 MSE = 4135, $p = .08$). This effect was somewhat surprising. The effect is in the opposite direction of that predicted by DM, which predicted a slow down in the 'Some N' condition. The UM predicted no difference between these two conditions in this region. UM could provide a post hoc account for the speed up on 'them were/did'. That is, if the specification of 'some' is achieved in the 'Some N' condition when the reader encounters the word 'all', then the information in this subsequent 'them were/did' region simply confirms an interpretation that has already been made and the reader can pass swiftly through this region.

Relevance theorists provide another possible interpretation. They have claimed that the processing of redundant information is costly, and that when a speaker uses repetition or redundancy, the speaker intends the hearer to derive extra contextual effects to off-set the extra costs of processing the repetition or redundancy. When one uses a plural definite description such as 'the books', one implies/presupposes that one is talking about the totality of some contextually specified group of books. Thus there is a felt redundancy in the sentence pair 'The books had color pictures. In fact all of them did…'. Hence there could be increased processing time in the cancellation sentence in the 'The N' condition. This would

presumably be localized in the ‘them were/did’ region, as that is the point at which the redundancy manifests itself, since that is the point at which the reader knows that the two sentences are talking about the same things. Under this account it is not that processing is speeded up in the ‘Some N’ condition but rather that it is slowed down in the ‘The N’ condition. The data do not discriminate between these two interpretations. But it is important to note that neither explanation favors the DM.

There was no difference in first pass time on ‘them were/did’ between the ‘Some N’ and ‘Many N’ conditions (Both F_1 and $F_2 < 1$). This appears to be consistent with DM. However, DM predicted that the overall reading patterns for the ‘Some N’ and ‘Many N’ would be similar to each other and different from the ‘The N’ condition. This prediction was not upheld. Readers spent more time on the word ‘all’ in the ‘Some N’ condition compared to the ‘The N’ condition. But there was no difference between ‘Many N’ and ‘The N’ in that region. The UM made no strong predictions regarding similarities or differences between these conditions.

Other regions of potential interest: DM predicted that there would be some evidence of increased processing effort reflecting the readers’ retrieval or construction of a new interpretation following the cancellation of the default GCI. There was no evidence of this in our data. Readers did not differ in time spent in the end-of-sentence region, or rereading the initial ‘Det N’ region as a function of determiner condition. This is consistent with the UM assumption that the specification is made incrementally as relevant information accrues and thus there is no need to reanalyze in the circumstances portrayed in our materials.

3.7 Discussion

Overall, the model best supported by our data is the UM, since it predicted a slow down on 'all' in the 'Some N' condition as compared to the 'The N' condition, and it predicted no overt reanalysis of the 'Det N' phrase. Moreover, it is able to give a reasonable explanation for the speed up in processing on the 'them were/did' phrase in the 'Some N' condition. But given that the results presented above constitute the first demonstration of these processing time effects and given that one might raise questions regarding the appropriateness of ‘The N’ as the control condition, we ran a second version of our experiment with a new control condition.

In this second version, processing of the ‘Some N’ sentence pairs from Experiment 1 was compared to processing patterns on sentence pairs that began with the quantifier phrase ‘At least some N’. This new version was intended as a case in which DM does not predict the triggering of the default GCI ‘Not all N’.

At the same time it does not involve a potential redundancy or repetition, as arguably the 'The N' condition does.[15] Thus readers saw sentences of the following sort in either its (a) or its (b) version:

(14)a. Some books had color pictures. In fact all of them did, which is why the teachers liked them.

b. At least some books had color pictures. In fact all of them did, which is why the teachers liked them.

24 participants from the University of South Carolina community (different from those who participated in Experiment 1) were recruited for this version. They either received one experimental credit in a psychology course or were paid $5 an hour for their time. All participants had normal vision and were native speakers of English. The procedure was the same as for Experiment 1. Again, a yes/no question (without feedback) followed 20% of the items, and all participants performed at 80% or better on comprehension questions. The analysis and regions of interest were the same as for Experiment 1.

Table 3 below presents our results. The results of most interest were again to be found in the 'all' and 'them were/did' regions in the cancellation sentences.

Table 3: Mean reading times (in msec) for 2 critical regions in the "cancellation" sentence in the 'Some N' and 'At least some N' conditions

	Gaze Duration on 'all'	First Pass Time on 'them were/did'
Some N	264	303
At least some N	249	325

The regions 'all' and 'them were/did': The most striking thing to note in Table 3 is the similarity in the pattern of data obtained here and in Experiment 1. Once again readers spent more initial processing time on 'all' following 'Some N' than in the control condition, in this case 'At least some N' ($F_1(1,23) = 3.57$, MSE = 766, $p < .07$; $F_2(1,62) = 3.83$ MSE = 1019, $p = .05$). In addition, readers spent less time in the region that forces the some and possibly all interpretation (viz. on the words 'them were/did') in the 'Some N' condition. But this result was not statistically significant ($F_1(1,23) = 1.07$, MSE = 5250, $p < .31$; $F_2 < 1$). Although the pattern and magnitude of these results replicates our previous results, the data were more variable and thus the statistical support for the results of this experiment alone is not as strong. As with the previous experiment we also looked at second pass time and regressions in to these regions and at second pass times on the 'Det N' phrase. No differences between the two conditions were found.

[15] Thanks to Kent Bach for the suggestion of this control.

If we assume that the predictions of the various models are similar to those displayed in Table 1, then again the UM is best supported. It predicts a slow down on the word 'all' in the 'Some N' condition as compared with the control 'At least some N' condition. The lack of evidence of any differential reanalysis effects between the two conditions is also consistent with the UM account. The numerical differences in the first pass time in the 'them were/did' region suggest that the pattern observed in Experiment 1 was not because readers were slowing down in the 'The N' condition due to a felt redundancy. However, we would not want to make any strong claims in this regard since the difference observed in the second experiment was not statistically reliable.

4. General Discussion and Conclusions

We have presented two very basic models based on two contrasting theoretical explanations of GCIs and data from two eye movement experiments that test predictions generated from the models. As far as we are aware, there is no previous published research using the eye monitoring methodology to study the way in which sentences that give rise to GCIs are processed.

Our results are most consistent with the UM. However, while our results pose problems for the DM as we conceived it, this model could be modified to accommodate the results we obtained. For instance, a version of the Default Model in which readers abandon the default in the face of potentially conflicting information (the word 'all' in our materials) rather than waiting until forced to do so (the 'them were/did' region) would make predictions that are largely compatible with the data that we obtained. Unfortunately, this is accomplished at the expense of compromising much of the utility ascribed to the notion of defaults in the first place. If the default can be dislodged even before the reader has conclusive evidence against it, it is unclear what the utility is of deriving it in the first place. Moreover, this Early Cancellation version of DM still does not account for the pattern of data observed in the 'them were/did' region, and it faces the problem that we found no evidence that readers engage in the sort of reanalysis predicted by the DM. Perhaps the more fundamental point to be made here is that whichever theoretical view one favors, these experiments provide new evidence of how sentences that give rise to GCIs are processed over time. We believe that this level of description can be put to good use to advance theory development in this area.

There has of course been other experimental work directed to studying people's comprehension of scalar and other generalized implicatures. For instance, Noveck (2001) and Chierchia et al. (2001) have investigated children's understanding of scalar implicatures. Also, several of the contributors to this volume, including Noveck, Guasti, and Politzer, report experimental investigations of scalar and clausal implicatures. These studies rely on judgment tasks in which people are asked to respond in some way to sentences that give rise to implicatures (e.g., people are asked to agree or disagree with these sentences, given some prior story context). The data produced by such experiments is very valuable. It can provide

evidence that people are aware of implicatures, but it cannot tell us at what point in the comprehension process implicatures are derived or whether implicatures are derived via default inferences. We have argued that what is needed in order to investigate the competing claims of DM and UM is data that is gathered during on-line processing of such sentences. Monitoring people's eye movements during reading can provide this sort of window onto on-line processing.

The eye monitoring methodology has been used successfully to study a variety of phenomena, including processes as diverse as lexical ambiguity resolution, sentence parsing, inferencing during reading, comprehension of metaphors and jokes, and comprehension of metonymies. There is every reason to think that it can also be fruitfully applied to investigate the processing of GCIs. It may even be that work applying this methodology in other areas can be mined for insights about how to deal with the processing of GCIs. For instance, Frisson and Pickering (1999) and Pickering & Frisson (2001) have used this methodology to study the processing of metonymies. The sorts of metonymy that they have been interested in are cases such as:

(15) The mayor was invited to talk to the school.

(16) My great-great-grandmother loved to read Dickens.

(15) is an example of a place-for-institution metonymy, whereas (16) is a producer-for-product metonymy. Frisson & Pickering have shown in a series of eye-tracking experiments that people have no difficulties with such metonymies. These are processed just as quickly as cases in which the noun phrases in question are used 'literally', such as in 'The demolition crew tore down the school' or 'My great-great-grandmother loved to see Dickens'. At least, this is the case when there is some familiar type of institution associated with the critical noun phrase, or when the person whose name is used is famously associated with the product referred to. On the other hand, they found that novel metonymies such as 'My great-great-grandmother loved to read Needham', when they are presented in neutral contexts, cause people processing difficulties compared with 'literal' uses and familiar metonymies. However, if in the preceding context it is mentioned that Needham is an author, readers experience no difficulties with a metonymical use of the name 'Needham', such as 'My great-great-grandmother loved to read Needham'.

Frisson and Pickering interpret their results as lending support to what they call the Underspecification Model. They argue that noun phrases of the sort they have studied express semantically underspecified meanings. But these underspecified meanings point to more specific meanings. When a word is first encountered, only the underspecified meaning is accessed initially. Which specific meaning (if any) is accessed will depend on the context. They compare their view to the one defended by Barton & Sanford (1993), according to which readers initially engage only in shallow

semantic processing, in which only the “core meaning” of an expression is accessed. Further meaning refinements are possible and which ones are made will depend on the context. Sometimes a reader may go no further than the core meaning, if it seems to fit with the discourse context (and Barton & Sanford use this fact to explain why the so-called Moses Illusion tricks people).

Although we are skeptical about the "core meaning" interpretation of semantic underspecification (see section 2.2 above), we think Frisson and Pickering's work is highly related to our own. They make a convincing case for the claim that only a methodology such as eye tracking can help decide between the Underspecification Model and some of the rival models they identify. They argue that their eye-tracking results count against any theory that argues that one particular meaning is the default. This applies, whether one privileges the 'literal' meaning or one of the metonymical ones. Frisson and Pickering found that at the earliest moments of processing there is no difference in reading times between phrases that require a 'literal' interpretation and ones requiring a metonymical reading. They relied on measures such as first fixation and first pass time, which give a picture of early processing. These are opposed to measures such as total time or number of regressions, which arguably give a picture of processing during the stage at which interpretations must be integrated into a discourse level representation. See Rayner, Sereno, Morris, Schmauder & Clifton (1989).[16]

Implicating something in a generalized way is clearly not the same thing as using a word or phrase metonymically, but metonymy is related to the sort of polysemy that has been of concern to cognitive linguistics such as Pustejovsky (1995) and Tyler & Evans (2001). The notion of semantic underspecification is central to the account these cognitive linguists have given of polysemy, and they have done a lot to clarify the notion of semantic underspecification. The notion of semantic underspecification plays a crucial role in our own work on GCIs, given that we are concerned to explore the competing pictures painted by the UM and the DM. Thus we think that work in cognitive linguistics and work in psycholinguistics can be fruitfully brought together, and the study of the processes involved in the derivation of GCIs is a good place at which to bring these fields together.

In this paper we hope to have done three things. We hope to have convinced readers (i) that it is worth investigating the way in which people process sentences of the sort that Grice thought give rise to GCIs, (ii) that there are at least two competing models that give an initially plausible picture of what this

[16] For a view opposed to Frisson & Pickering’s, see Giora (In press). Giora argues that with polysemies there is no need to select between the alternative meanings, and so all possible meanings can be accessed simultaneously, without any impact on processing. Thus she is skeptical that Frisson & Pickering have unequivocal evidence for their underspecification model.

processing is like, and (iii) that processing data from eye monitoring can give us the sort of evidence that is needed to test and refine these models.[17]

[17] We would like to acknowledge the assistance of Carter Henderson, Sachiko Matsumoto, Rihana Williams, Beth Myers, Matt Traxler and Johnny Hancock. This paper was first presented to the Experimental Pragmatics Workshop held in Lyon, France in May 2001. We would like to thank those in the audience who gave us advice. Special thanks to Dan Sperber and Ira Noveck, both in their role as hosts of the Experimental Pragmatics Workshop, and as editors of this volume, and also to Rachel Giora, who read the penultimate version of this paper and gave us extensive and helpful feedback. This material is based upon work supported by the National Science Foundation under Grant No. 0080929.

References

Atlas, J. D., & Levinson, S. C. (1981). It-clefts, informativeness, and logical form: Radical pragmatics (Revised standard version). In P. Cole (Ed.), Radical Pragmatics (pp. 1-61). New York: Academic Press.

Bach, K. (1994a). Conversational impliciture. Mind & Language, 9, 124-162.

Bach, K. (1994b). Semantic slack: What is said and more. In S. L. Tsohatzidis (Ed.), Foundations of Speech Act Theory: Philosophical and Linguistic Perspectives (pp. 267-291). London: Routledge.

Bach, K. (1995). Standardization vs. conventionalization. Linguistics and Philosophy, 18, 677-686.

Bach, K. (1998). Standardization revisited. In A. Kasher (Ed.), Pragmatics: Critical Assessment. London: Routledge.

Bach, K. (1999). The semantics-pragmatics distinction: What it is and why it matters. In K. Turner (Ed.), The Semantics/Pragmatics Interface from Different Points of View (pp. 65-84). Oxford: Elsevier.

Barton, S., & Sanford, A. J. (1993). A case-study of pragmatic anomaly-detection: Relevance-driven cohesion patterns. Memory & Cognition, 21, 477-487.

Bezuidenhout, A. L., & Cutting, J. C. (2001). Literal meaning, minimal propositions and pragmatic processing. Journal of Pragmatics.

Carston, R. (1988). Implicature, explicature and truth-theoretic semantics. In R. Kempson (Ed.), Mental Representations: The Interface between Language and Reality (pp. 155-182). Cambridge: Cambridge University Press.

Carston, R. (1993). Conjunction, explanation and relevance. Lingua, 90, 27-48.

Carston, R. (1995). Quantity maxims and generalized implicature. Lingua, 96, 213-244.

Carston, R. (1997). Enrichment and loosening: Complementary processes in deriving the proposition expressed? Linguistische Berichte, 103-127.

Carston, R. (1998a). Informativeness, relevance and scalar implicature. In R. Carston & S. Uchida (Eds.), Relevance Theory: Applications and Implications (pp. 179-236). Amsterdam: John Benjamins.

Carston, R. (1998b). Pragmatics and the Explicit- Implicit Distinction. Unpublished Ph.D.dissertation, University College London, London.

Chierchia, G., Crain, S., Guasti, T., Gualmini, A., & Meroni, L. (2001). The acquisition of disjunction: Evidence for a grammatical view of scalar implicatures. Paper presented at the 25th Annual Boston University Conference on Language Development.

Coulson, S. (2001). Semantic Leaps: Frame-Shifting and Conceptual Blending in Meaning Construction. Cambridge: Cambridge University Press.

Fauconnier, G. (1985). Mental Spaces: Aspects of Meaning Construction in Natural Language. Cambridge, MA: MIT Press.

Fauconnier, G. (1997). Mappings in Thought and Language. Cambridge: Cambridge University Press.

Folk, J. R., & Morris, R. K. (1995). Multiple lexical codes in reading: Evidence from eye movements, naming time, and oral reading. Journal of Experimental Psychology: Learning, Memory, and Cognition, 21, 1412-1429.

Frisson, S., & Pickering, M. J. (1999). The processing of metonymy: Evidence from eye movements. Journal of Experimental Psychology: Learning, Memory and Cognition, 25, 1366-1383.

Gazdar, G. (1979). Pragmatics: Implicature, Presupposition, and Logical Form. New York: Academic Press.

Gibbs, R. W., & Moise, J. F. (1997). Pragmatics in understanding what is said. Cognition, 62, 51-74.

Giora, R. (In press). On Our Mind: Salience, Context, and Figurative Language. New York: Oxford University Press.

Grice, P. (1989). Studies in the Way of Words. Cambridge, MA: Harvard University Press.

Hirschberg, J. (1991). A Theory of Scalar Implicature. New York: Garland Publishing, Inc.

Horn, L. 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 (pp. 11-42). Washington, D.C.: Georgetown University Press.

Horn, L. R. (1989). A Natural History of Negation. Chicago: University of Chicago Press.

Levinson, S. C. (1983). Pragmatics. Cambridge: Cambridge University Press.

Levinson, S. C. (1987a). Implicature explicated? Behavioral and Brain Sciences, 10, 722-723.

Levinson, S. C. (1987b). Minimization and conversational inference. In J. Verschueren & M. Bertuccelli-Papi (Eds.), The Pragmatic Perspective (pp. 61-129). Amsterdam: John Benjamins.

Levinson, S. C. (1995). Three levels of meaning. In F. R. Palmer (Ed.), Grammar and Meaning (pp. 90-115). Cambridge: Cambridge University Press.

Levinson, S. C. (2000). Presumptive Meanings: The Theory of Generalized Conversational Implicature. Cambridge, MA: MIT Press.

Nicolle, S., & Clark, B. (1999). Experimental pragmatics and what is said: A response to Gibbs and Moise. Cognition, 69, 337-354.

Noveck, I. A. (2001). When children are more logical than adults: Experimental investigations of scalar implicature. Cognition, 78, 165-188.

Pickering, M. J., & Frisson, S. (2001). Why reading Dickens is easy (and reading Needham is hard): Contrasting familiarity and figurativeness in language comprehension.

Poynor, D.V. & Morris, R.K. (under review) Inferred Goals in Narratives: Evidence from self-paced reading, recall and eye movements. Journal of Experimental Psychology: Learning, Memory and Cognition.

Pustejovsky, J. (1995). The Generative Lexicon. Cambridge, MA, USA: MIT Press.

Pustejovsky, J. (1998). The Semantics of Lexical Underspecification. Folia Linguistica, 32(3-4), 323-347.

Rayner, K., & Morris, R. K. (1990). Do eye movements reflect higher order processes in reading? In R. Groner & G. d'Ydewalle & R. Parham (Eds.), From Eye to Mind (pp. 179-191): North Holland Press.

Recanati, F. (1991). The pragmatics of what is said. In S. Davis (Ed.), Pragmatics: A Reader (pp. 97-120). New York: Oxford University Press.

Recanati, F. (1993). Direct Reference: From Language to Thought. Oxford: Blackwell.

Recanati, F. (1995). The alleged priority of literal interpretation. Cognitive Science, 19, 207-232.

Sperber, D., & Wilson, D. (1986). Relevance: Communication and Cognition. Oxford: Blackwell.

Tyler, A., & Evans, V. (2001). Reconsidering prepositional polysemy networks: The case of over. Language, 77(4), 724-765.

van Deemter, K., & Peters, S. (Eds.). (1996). Semantic Ambiguity and Underspecification. Stanford: CSLI Publications.