Chapter 14

Pragmatic inferences related to logical terms

Ira A. Noveck

1. Introduction

Paul Grice was concerned with the way logical terms such as *some*, *or*, and *and* take on extralogical meanings in conversational contexts. To take one example, Grice (1989) described *or* as having a weak word meaning identical to formal logic's inclusive disjunction (which is false only in the case where both disjuncts are) but as conveying in conversation a stronger speaker's meaning corresponding to the exclusive disjunction (which is false in the case where both disjuncts are false or where both are true). Grice used the term *implicature* to describe the pragmatic inference linking word meanings to speaker's meanings and laid the foundations for nearly all of the linguistic-pragmatic studies found in this volume. [1]

The idea of submitting Grice's hyoptheses to experimental investigations is extremely attractive. But what would testable processing predictions look like precisely? Grice assumed that hearers expect speakers, in producing an utterance, to obey a set of maxims following from a general co-operative principle. When their initial interpretation of an utterance fails to confirm that the speaker has obeyed the maxims, or at least the

[1] The pragmatic literature has fine-tuned the notion of implicature (Bach, 1994, Sperber & Wilson, 1986/1995) making it a confusing term for describing the inferences to be discussed here. I will refer to the derived extra-logical meanings generically as pragmatic inferences or specifically as either a scalar inference (e.g.when *but not both* is derived from *or*) or as pragmatic enrichment (when *and* is treated as *and then*).

co-operative principle, hearers derive implicatures so as to reconcile the overall interpretation of the utterance with their expectation. How does this work exactly? In general, one would expect that the derivation of an implicature should involved extra processing, but Grice does not provide enough detail. Moreover Grice suggests that some implicatures -- his so-called 'generalized conversational implicatures' linked in particular to words such as *some*, *and*, or *or* -- are derived by default, and may be contextually cancelled. For these generalized implicatures, it is their cancellation that should involve extra processing. More recent theoretical work in pragmatics has both sharpened Grice's ideas and provided alternative accounts, bringing us closer to formulating experimentally testable hypotheses.

In what follows, I briefly review the proposal from neo-Griceans, focusing on Levinson (2000), before turning to Relevance Theory (Sperber and Wilson, 1986/1995). I will then show how the developmental psychological literature had been investigating implicatures over 20 years ago, albeit unwittingly. The rest of the chapter describes several experiments that I and colleagues of mine have been carrying out in our lab, with children and adults, in order to better understand how pragmatic inferences linked to logical terms are generated.

2. Two Post-Gricean approaches

According to neo-Griceans, such as (Horn, 1973) and Levinson (Levinson, 1983; 2000), the scalar inference illustrated by Noemi's response in (1) is a case that works on *terms* that are relatively weak:

(1) [*Knock at the door*]

Isaac: Is that Mama and Papa?

Noemi: It's Mama or Papa.

Noemi's choice of a weak term *or* implies the rejection of the stronger term *and*. More specifically, the connectives *or* and *and* may be viewed as elements of a scale (<or, and>), where *and* constitutes the more informative element (since *p and q* entails *p or q*). In the event that a speaker chooses to utter a disjunctive sentence, *p or q*, the hearer will take it as suggesting that the speaker either has no evidence that a stronger element in the scale, i.e. *p and q*, holds or that she perhaps has evidence that it does not hold. Presuming that the speaker is cooperative and well informed the hearer will tend to infer that it is not the case that *p and q* hold, thereby interpreting the disjunction as exclusive. This neo-Gricean analysis can be extended to other logical terms. For example, if a speaker uses the weak quantifier *some* (as in *some triangles are equilateral*), it implies that the stronger quantifier *all* is not appropriate. If one uses the modal *might* (as in *Bill might be in the office*), it implies that the speaker had reason *not* to say the stronger-sounding *must* (as in *Bill must be in the office*). However, the neo-Gricean account is not limited to logical terms and it has been applied to a host of scales initially described by Horn (for a review, see Levinson, 1983; 2000). Other possible scales are frequency (where the use of *sometimes* excludes *always*) and epistemic status (where the weaker *think* implies that it is not the case that *know*). In each case, scales range from less to more informative and the speaker's use of a less-informative term implies the exclusion of a more-informative one.

Levinson (2000) motivates his most recent account by pointing out how "the remarkably slow transmission rate of human speech," leads to a "bottleneck in the efficiency of human communication" (page 28). His proposal for surmounting the bottleneck is to profit from one's relatively high speed of comprehension in order to treat pragmatically enriched meanings of certain terms as a "default" or as a "preferred" meaning. These preferred meanings are put in place as a result of heuristics. For example, scalars are considered by Levinson to result from a Q-heuristic, dictating that "What isn't said isn't (the case)." It is named *Q* because it is directly related to Grice's (Grice, 1989) first maxim of quantity: *Make your utterance as informative as is required*. In other words, Levinson assumes that scalar inferences are general and automatic. When one hears a weak scalar term like *or, some, might* etc. the default assumption is that the speaker knows that a stronger term from the same scale is not warranted or that she does not have enough information to know whether the stronger term is called for. This means that relatively weak terms prompts the inference by default -- *or* becomes *or but not both*, *some* becomes *some but not all* etc. Scalar inferences by default can be cancelled, but the very idea of cancellation (as opposed to, for instance inhibition) implies that it must occur *subsequent* to the production of the inference.

The other account comes from Relevance Theory (Sperber & Wilson, 1986/1995), which assumes that the interpretation of an utterance can be inferentially enriched in order to better capture the speaker's intention, but such pragmatic enrichment is not achieved through context-insensitive default inferences triggered by the mere presence of a weak scalar term. According to Relevance Theory, the so-called

scalar inferences are ordinary pragmatic inferences drawn by hearers in order to arrive at an interpretation of an utterance that meets their expectations of relevance. How far the hearer goes in constructing an utterance's interpretation is governed by considerations of effect and effort; hearers expect the intended interpretation to provide satisfactory effect for minimal effort.

A non-enriched interpretation of a scalar term (the one that more closely coincides with the word's meaning) can often lead to a relevant-enough interpretation of the utterance in which it occurs. Consider *Some monkeys like bananas*. This utterance with an interpretation of *Some* that remains in its weaker form (this can be glossed as *Some and possibly all monkeys like bananas*) can suffice for the hearer and not require further pragmatic enrichment. In contrast, the potential to derive a scalar inference comes into play when an addressee has higher expectations of relevance. A scalar inference could well be drawn by a hearer in an effort to make an utterance more informative and thereby more relevant. Common inferences like scalars are inferences that optionally play a role in such enrichment; they are not steadfastly linked to the words that could prompt them. When a scalar inference takes place and renders an underinformative utterance more informative, it ought (all things being equal) to involve extra effort.

One can better appreciate the two accounts by taking an arbitrary utterance (2) and comparing the linguistically encoded meaning (3a) and the meaning inferred by way of scalar inference (3b):

(2) Some X are Y.

(3) a. Some and possibly all X are Y (logical interpretation).

b. Some but not all X are Y (pragmatic interpretation).

Note that (3a) is less informative than (3b) because the former is compatible with four possibilities: i) X is a subset of Y, ii) Y is a subset of X, iii) X and Y overlap, and iv) X and Y coincide whereas interpretation (3b) is compatible only with possibilities (ii) and (iii) and therefore reduces the range of possible states of affairs described by the used of *some*. According to Levinson, the interpretation in (3b) is prepotently adopted through the Q-heuristic. This becomes the default meaning unless something specific in the context leads one to cancel the inference giving rise to (3b) and to then adopt the reading in (3a).

According to Relevance Theory, a hearer starts with the most accessible interpretation, which, in the absence of contextual cues, is provided by the plain linguistic meaning of a word such as "some", as in (3a); if that reading is satisfactory to the hearer, she will adopt it. However, if interpretation (3a) fails to meet the hearer's expectation of relevance, she may enrich it and adopt interpretation (3b) instead. Given that (3b) arrives by way of a supplementary step (an inference), there is a cost involved (i.e. cognitive effort). This amounts to deeper processing, but at a cost.

3. Classic experimental findings

As this volume exemplifies, only recently has there been a concerted effort to tackle linguistic-pragmatic issues experimentally. However, it is important to point out that there are some classic reasoning studies (Braine & Rumain, 1981; Evans & Newstead, 1980; Paris, 1973; Smith, 1980; Sternberg, 1979) that serve as a prelude for

work concerning the present issues. These prior studies yield two kinds of results. One is that when adult participants are presented weak scalar utterances in contexts in which a stronger reading is possible but not necessary they are often equivocal between logical and pragmatic interpretations. For example, consider two studies on disjunction: one conducted only with adults and a developmental one that included adults. Evans and Newstead (1980; Exp. 2) presented participants with a rule about the presence of a letter and a number on a screen: "Either there is a P or a 4" along with exemplars (such as a P with a 4, a P with a 9, or a Q with a 4 etc.). When the two exemplars are P and 4 (i.e. presented as "P 4"), the authors report that 57% of participants reply that such a rule is true of this combination (and given that the task required a forced-choice, 43% said that it was false). Similarly, Paris's (1973) developmental study showed that 67.5% of adults respond true when presented two images, e.g. a boy next to a bicycle and a monkey in a tree, and told to evaluate "Either the boy is next to the bicycle or the monkey is in the tree" (and 75% of adults respond true if the formulation excludes *Either* as in "The boy is riding the bicycle or the monkey is in the tree").

The other, more remarkable result is that, provided identical situations, children are more likely than adults to provide logically correct responses. In the Paris (1973) study, upwards of 90% of 8-year-old children accept (as true) cases where both disjuncts were true; moreover, these children were significantly more likely than adults to respond *True* to these cases and the developmental trend (ranging across five ages) is linear. This developmental effect appears robust since it is found elsewhere in the literature (Braine and Rumain, 1981; Sternberg, 1979; see also Smith, 1980). Here is what

Sternberg (1979, p. 492) plainly said after coming across the same result that Paris reported:

> The data show an interesting interaction between age and interpretation of *or*…children at the lowest grade level use the inclusive interpretation of *or* in preference to the exclusive interpretation…At the higher grade levels, children show a strong tendency to use the exclusive interpretation in preference to the inclusive interpretation.

What is absent in these papers is an explanation for this effect. Although researchers recognized its curious nature, they were generally mystified by it. After all, it was rare to find children behaving more logically than adults. Viewed through the prism of pragmatics, however, the effect becomes obvious: the linguistically encoded meanings of weak terms (like *or*, *some*, and *might*) are compatible with minimal interpretations of underinformative items while pragmatic inferences increase with age. That is, the minimal interpretation for each of these terms is compatible with a logical one (e.g. when someone says *or*, it means that at least one disjunct is true, when someone says *some*, it means that at least one of several quantified objects is the case etc.). Pragmatic enrichments provide for the adult responses. That is, adults are more likely to apply pragmatic principles than children and thus enrich the interpretation of *or* by adding *but not both*. This pragmatic insight could readily solve a small mystery in the developmental literature. The next section shows how just how general this effect is.

4. Establishing the developmental-pragmatic effect

Once appreciated as a pragmatic effect, understanding it even better has played a role in creating a small cottage industry of experiments in my lab and others'. My lab undertook a series of experiments that employed two classes of weak scalar utterances, modals and existential quantifiers (Noveck, 2001). For modals, the critical test item was *there might be a parrot* when something like *there must be a parrot* would be more appropriate. For quantifiers, the test items were statements like S*ome elephants have trunks* (and we know that *All elephants have trunks*). The scenarios from this paper along with its main results are presented in the next two sections.

4.1 The modal *might*

Consider a reasoning task involving three boxes. One is open and has a toy parrot and a toy bear in it (the Parrot+Bear Box), the second is open and has only a parrot (the Parrot-only Box), and the third stays covered (Box C). Participants are told that Box C has the same content as either the Parrot+Bear Box or the Parrot-only Box. A puppet presents eight statements and it is the participant's task to say whether the puppet's claim is right or not. The critical statement that allows us to study the pragmatic inference is *There might be a parrot in the box* when the evidence shows that there *must* be a parrot. On the one hand, if the participant adopts an explicit, logical interpretation of *Might* (where *Might* is compatible with *Must*), one would expect an affirmative reply ("the puppet is right"). On the other hand, if the participant adopts a pragmatic,

restrictive interpretation for *Might* (where *Might* is not compatible with *Must*) one would expect a negative reply ("the puppet is wrong") or at least some equivocation. Seven-year-olds' rate of logical interpretations with respect to *There might be a parrot in the box* (80%) is intriguing not only because they respond at rates that are significantly above chance levels but because they do so at a rate that is significantly higher than that of the adults' (35%), which resembles chance levels.

4.2 Existential quantifiers

In another experiment investigating quantified statements, 8-year-old and 10-year-old children and adults were confronted with various statements including a set that can be exemplified with *Some elephants have trunks*. These kinds of utterances have the potential to generate scalar inferences of the sort *Not all elephants have trunks* which leads to a contradiction with one's stereotypical knowledge about elephants. Two features of these studies are worth pointing out. One is that there were 6 kinds of statements overall based on there being Existential and Universal Quantifiers (*Some* and *All*) and three kinds of relations: A) Absurd (e.g. Some chairs tell time/All crows have radios), B) Appropriate (e.g. Some houses have bricks/All elephants have trunks), and Inappropriate (e.g. Some giraffes have long necks/All dogs have spots). It is the *Inappropriate* condition for *Some* that presents us with the infelicitous, underinformative case. The other important feature is that the test was conducted under double blind conditions. The experimenter was simply told to present the utterances in an even tone of voice before soliciting an Agree/Disagree response (and the experimenter later told me that she thought the absurd statements were the critical ones). Aside from near

perfect responses to the five indisputable statements, the results showed that all the children were very likely to agree (at rates at or above 85%) with the Underinformative statements and significantly more so than adults, who were split in their responses (41% of participants consistently agreed with the Underinformative statements).

5. Understanding the developmental-pragmatic effect even better

I have presented just two examples of new studies investigating this developmental-pragmatic effect that was first unearthed in pre-existing literature. Others have reported similar findings, making this a robust effect (see Guasti et al. in preparation, Chierchia et al., this volume, Chapter 13; Papafragou & Musolino, 2003). When a relatively weak term is used in scenarios where a stronger term is justified, younger children are typically more likely than adults to find the utterance acceptable.

Several issues remain. Some colleagues have reacted to my data as if I mean to say that children are pragmatically delayed at young ages (Meroni, Gualmini, & Crain, 2001). This is very far from what is claimed. In fact, it was anticipated that "one would find the same effect among even younger children if a task were made easy" enough (Noveck, 2001, page 184). Another issue concerns the relevance of this effect to resolving the debate between the two opposing accounts of scalar inferences compared earlier. These developmental findings do not favor one account over another because both could explain it. From the Default inference perspective, it could be claimed that scalar inferences become automatic with age and that our results are simply revealing how such inference-making matures. In contrast, Relevance Theory would suggest that children and adults use the same comprehension mechanisms but that greater cognitive

resources are available for adults, which in turn encourages them to draw out more pragmatic inferences.

One way to address both of these issues is by finding a link between scalar-inference production and task complexity. If one assumes that cognitive effort is indeed a critical factor in such inference-making, a simpler task ought to make its production more likely (for similar arguments, see Noveck, Chierchia, Chevaux, Guelminger, & Sylvestre, 2002). Such a finding would have to be considered favorable to Relevance Theory because it would indicate that relative effort of utterance interpretation in *context* is critical for implicature production. If scalar implicatures were automatic and linked to particular words, task complexity ought not to matter. This is what Nausicaa Pouscoulous, Guy Politzer, Ann Bastide and I aimed to show in a series of studies conducted with much younger children. In these studies, we show that, even in the simplified scenarios, one finds that children become progressively more pragmatic with age.

5.1 Existential quantifiers again

We employed a reasoning scenario that tested quantifier comprehension but that (unlike the above study on *Some*) did not require long-term knowledge. In our first experiment, we presented a standard paradigm that placed four cardboard boxes in front of participants with different plastic animals placed in and around the boxes. The 9-year-old and adult participants were asked whether they agreed with a puppet that made statements about the scenario. Among numerous control items was the critical test statement – "Some turtles are in the boxes" (*Certaines tortues sont dans les boîtes*) --

when all of the presented turtles were in the boxes. The responses to these statements are then used as a dependent variable to assess whether participants make the scalar inference. If participants were to make the scalar inference, they would disagree with the puppet (because *all* the turtles are in the boxes), whereas if they treated the word *some* in a logical way they would agree with the statement. The results confirmed the previous findings (92% of children responded logically whereas 47% of adults did). This is the scenario that was modified so that younger children could be tested.

We made three changes which we believed would reduce the effort necessary to perform the task, thus encouraging the pragmatic responses in children. First, we used the French word *quelques* instead of *certains*. Although both words mean *some*, teachers indicated that children were more comfortable with the former. Second, the presentation concerned only tokens that were in boxes; there were no animals and these tokens were not left strewn around outside the boxes (as was the case for the animals in the standard paradigm). All the statements concerned the contents of the boxes. Third, we asked participants to perform an action on the basis of the puppet's instructions rather than making a judgment on the validity of the puppet's statements. Participants were asked to fulfill a wish made by a puppet concerning the items in the boxes. For example, the puppet would say "I would like all of the boxes to have a token", and the participant would have to determine whether or not they should alter the scenario in order to comply with the wish. Of course, the utterance of interest arose when the puppet said *Je voudrais que quelques boites contiennent un jeton (I would like some of the boxes to have a token*) when all of the boxes already contained a token. If participants believed that *some* was compatible with *all* they ought to leave the boxes unchanged (which is

actually difficult to do given that a request implies that a change is called for); otherwise they could remove some of the tokens.

When very young children (4 year olds, 5 year olds, and 7 year olds) are presented with a scenario having four boxes and each with a token, one finds many more responses indicating scalar inference generation than in the prior experiments. Only 32% of the 4-year-olds and 27% of the 5-year-olds gave a response that indicated that they chose the logical response (i.e. they left the boxes untouched); even fewer seven year olds did so (17%). Scalar-inference making is more apparent in the children even as the developmental ordering remains marked. Pragmatic inference-making is affected by task ease (and appears to continue to be affected by the level of sophistication on the part of the hearer). This is evidence that the ability to draw implicatures is not uniquely linked to maturity. By making a task easy enough, we have encouraged children to apply their available resources to draw the scalar inference.

5.2 Pragmatically enriching "and"

Does this developmental effect extend to non-scalar cases? Here, we focus on a pragmatic inference linked to *and* which prompts the same developmental tendency as those described for scalars. To appreciate the pragmatic enrichment in question, consider the two conjunctive utterances in (4).

(4) a. Mary got married and got pregnant.

b. Mary got pregnant and got married.

The two are equivalent from a logical point of view because they both contain the same two components, i.e. *P & Q = Q & P*. However, the sequence of the conjuncts in each of the two utterances conveys two very different sets of implications. Whereas it would be considered a normal occurrence to hear about someone getting married before getting pregnant (4a); in some parts of the world, it would be considered scandalous to get pregnant before getting married (4b). Without the implicit sequential interpretation, the statement in (4b) would not seem scandalous.

The pragmatic inference linked to *and* prompts the same debate as the one for scalars.[2] On the one hand, Levinson (Levinson, 1983; 2000) argues that interlocutors "buttress" conjunctions by interpreting *and* to mean *and then* and that they do this by default. On the other, Carston, (1996; 2002) points out that there are host of ways in which *and* can be enriched and argues that none dominate, that context determines which pragmatic enrichment to make and that there is nothing automatic about it. Below, are just five of the kinds of implications that a conjunctive utterance can convey:

(2) a. *Contrast:* It's autumn in the U.S. and it's spring in Chile.

b. *Sequential*: She took the scalpel and made the incision.

c. *Containment*: We spent the day in town and went to Macy's.

d. *Causal*: She shot him in the head and he died instantly.

---

[2] The linguistic intuition that the pragmatic enrichment of *and* is comparable to scalar inference making is not universally shared. Recanati (2003) assumes that the pragmatic enrichment of *and* is sublocutionary, making it (in contrast to scalars) not as readily available to consciousness.

e. *Indirect Causal*: He left her and she took to the bottle.

Carston's account is largely corroborated in two separate studies. Noveck and Chevaux (2001) presented seven- and ten-year-old children as well as adults a set of 12 stories about everyday events. Among these were four stories that ultimately presented a conjunctive sentence as a comprehension question. For half of the conjunctive comprehension questions, the two events in the question were presented in a sequence that respected the order in the stories and in the other half the sequence was inverted. To illustrate, consider the short story in (5) and its two kinds of follow-up questions in (6a) and (6b):

(5) While sitting on her couch, Julie was reading a comic book.

Suddenly, the phone rang.

She went out of the living room and ran to answer.

It was Isabelle who was inviting Julie to celebrate her birthday Saturday.

Since they were very good friends, Julie accepted the invitation.

(6a) Julie answered the phone and accepted an invitation?

(6b) Julie accepted an invitation and answered the phone?

Whereas the rates of agreement to (6a) are high and accurate for all participants, we found that linguistically competent children are less fussy than adults about sequence in conjunctive sentences. Roughly 85% of seven-year olds, 63% of ten-year-olds, and about 29% of the adults say "Yes" to (6b). This developmental curve resembles the one for scalars. Younger children are more likely to agree with a statement's minimal interpretation because they are less likely to pragmatically enrich the meaning of *and*.[3]

Another piece of evidence supporting Carston's account comes from processing studies that followed up on the paradigm above. Noveck, Chevaux & Bott (2004) enlarged the list of stories and presented them line-by-line on a screen to ten-year-olds and adults.[4] The dependent measures were the Yes/No response to the question as well as the response time it required. The developmental curve for the categorical responses was as before, although there was more pragmatic inferencing in general: The ten-year-olds were significantly more likely than adults to accept the inverted order as true (46% vs. 18%, respectively) and this was the only adult-child comparison to yield significant effects.

Of further interest were ten-year-olds' response times to the two sorts of responses because one could readily draw out processing predictions from the two opposing accounts. On the one hand, if the sequential (buttress) interpretation occurs by

[3] It is important to point out that only the ten-year-olds appeared adult-like in all other respects. However, the seven-year-olds tended to respond "Yes" to Control, Inverted-order questions that employed *and then* explicitly as its conjunction.

[4] The stories were highly similar to those in Noveck & Chevaux (2001), but were slightly modified for computer-presentation purposes and were not joined by filler items. There were nine stories altogether and each could be followed with any one of three kinds of comprehension questions: Order-preserved and Order-inversed as before, plus Order-preserved with a false second conjunct. This new type of category, is exemplified in (6c), which is a conclusion for the story shown in (5):

(6c) Julie answered the phone and declined the invitation.

default (or is automatic in some form) then "No" responses ought to be quickest and "Yes" responses ought to take at least as long because the latter response requires the cancellation of the buttress. On the other hand, a Relevance account predicts that those who respond "Yes" to the inverted-order questions ought to answer more quickly than those who answer "No" because the pragmatic enrichment occurs subsequent to minimal semantic treatments of *and*. The data fall in favour of Relevance accounts, which shows that children who answer "No" to the inverted-order questions require (on average) 2 seconds more than those who answer "Yes". Arguably this is not due to a negation bias, because the control question that prompts a correct "No" response led to fastest response times overall (see Footnote 4). Unfortunately, there were too few adult responses that would justify a similar analysis among them, but in the next section I describe adults' processing times in sentence verification tasks that investigate scalar inferences.

6. Adult studies

While the developmental studies are illuminating, it could be argued that they impose certain limitations. Perhaps children are categorically different from adults or perhaps their data are not as reliable as the adults'. Although I do not share these criticisms, support for Relevance claims would be stronger if the same sorts of effects could be found among adults. How could we investigate that? One way is to uncover the time course of pragmatic enrichments a bit more carefully. If one could provide evidence showing that pragmatic interpretations of scalars are the first to arise and that interpretations that require cancellations occur subsequently, then the default inference view would be supported. However, if one could show that minimal interpretations are

at the root of initial interpretations and that pragmatic interpretations arise only later, that would be further support for Relevance Theory.

6.1 Existential quantifiers

Lewis Bott and I (Bott & Noveck, 2003) set up three carefully controlled experiments which investigated the time course of statements like *Some monkeys are mammals*, which I have referred to as the Underinformative item. The design was drawn from Smith (1980) and Noveck (2001), but it benefited from several rigorous controls. I describe three of these.

First of all, our experiments included 6 kinds of sentences all of the form *[quantifier] A are B*, with two possible quantifiers (*Some* and *All*) and three kinds of set relationships between the As and the Bs -- one in which the As are a proper subsets of the Bs (*Some/All monkeys are mammals*), one in which the Bs are a proper subset of the As (*Some/All mammals are monkeys*) and a third where the As and the Bs form two disjoint sets (*Some/All monkeys are fish*). As can be seen in Table 14.1, these materials provide us with the Underinformative items as well as a range of controls that include both True and False items. This way we could compare both kinds of anticipated responses for the Underinformative items with many different controls.

Second, the experiment was designed so that any one of nine subordinate categories (e.g. monkeys) could be randomly joined in a quantified statement with any one of six superordinate categories (mammals, reptiles, fish, insects, shellfish, & birds). For example, one could insert “monkeys” to be part of *Some monkeys are mammals,*

Some mammals are monkeys, Some monkeys are fish, All monkeys are mammals, All mammals are monkeys, or *All monkeys are fish*. Each subordinate category was used just once and randomly per experiment which means every participant viewed a unique set of materials.

Finally, an on-line investigation allowed us to be creative with our presentation. One could present the sentences one word at a time or an entire sentence at a time. We could also require participants to hurry by putting time limits or encourage them to take their time. All told, this form of experimentation allows us to create various experimental contexts as we ask participants to respond to Underinformative items and their controls.

Before continuing the description of the experiments, it is important to point out that classic categorisation studies (much like the classic developmental studies mentioned earlier) did pay some attention to the infelicitous Underinformative items. However, most dealt with the bivocality of *Some* largely by sidestepping it. Response time experiments have generally instructed their participants to interpret *Some* in a strictly logical way (i.e. without the scalar inference). For example, Meyer (1970) told participants to treat *some* as meaning *some and possibly all* in a sentence verification task with sentences like *some pennies are coins*. To the best of our knowledge, there is only one psychological study to take an interest in the potentially conflicting interpretations of such underinformative sentences (Rips, 1975). Rips investigated how participants make category judgments by using sentence verification tasks with materials like *some congressmen are politicians*. He examined the effect of the quantifier interpretation by running two experiments, one in which participants were asked to treat

some as *some and possibly all* and another where they were asked to treat *some* as *some but not all*. This comparison demonstrated that the participants given the *some but not all* instructions in one Experiment responded more slowly than those given the *some and possibly all* instructions in another. Despite these indications, Rips modestly hedged when he concluded that "of the two meanings of Some, the informal meaning *may* be the more difficult to compute" (italics added). His reaction of course is not uncommon. Many colleagues share the intuition that the pragmatic interpretation seems more natural. In any case, this is an initial finding that goes in favour of the Relevance account.

Our studies picked up where Rips left off. In Experiment 1, we replicated Rips (1975, experiments 2 and 3) in one overarching procedure. Participants took part in two sessions. In one, participants were instructed to interpret the quantifier *some* to mean *some and possibly all*, which we refer to as the Logical condition and in the other, they were told to interpret *some* to mean *some but not all*, which we will refer to as the Pragmatic condition (and, of course, session order was counterbalanced). Central to our interests was the accuracy and the speed of response to the Underinformative sentences (e.g. *Some monkeys are mammals*) under the two conditions. According to the Default Inference account, a False response should be faster than a True response because the latter ought to occur as a result of the default inference's cancellation. In contrast, Relevance Theory would predict that a False response occurs more slowly than a true response because the False response would arise when Relevance conditions are applied more stringently, resulting in the production of the scalar inference.

The results showed that when participants were under instruction to, in effect, draw the scalar inference they required significantly more time to evaluate the underinformative sentences than when they were under instructions to provide a logical response. In fact, a response that entails the presence of a scalar inference (i.e. to say False to *Some monkeys are mammals*) is extraordinarily slow (i.e. slower than any other True or False response in the task). This will become a staple finding in this series of experiments.

The data also show that participants have greater difficulty providing the "correct" response when they are given Pragmatic instructions. Participants are accurate on approximately 85% of the Underinformative items under Logical instructions and accurate on about 60% of the Underinformative items under Pragmatic instructions. Moreover, the rate of correct responses for the Underinformative item under Logical instructions is comparable to the rates of correct responses to all of the control items, whereas the rate of "correct" responses for the Underinformative items under Pragmatic instructions appears exceptionally low. Thus, at least some of the extra time required under Pragmatic instructions is due to the processing requirements of making and maintaining the inference. This much confirms Rips's initial findings. There are no indications that turning *some* into *some but not all* is an effortless step.

Experiment 2 used the same paradigm but we provided neither explicit instructions nor feedback about the way to respond to the Underinformative sentences. Instead, we expected participants to answer equivocally to these types of sentences - some saying false and some true. This means that we should have two groups of responses: one in which the inference is drawn (what we call Pragmatic responses) and another where there

is no evidence of inference (Logical responses). We can therefore make a comparison between the two as we did in the previous experiment. Once again, if logical responses are made more quickly than pragmatic responses, we have evidence against a default system of inference. We can also use the control sentences to verify that under these more neutral conditions, responses which involve a pragmatic inference require more time than responses that do not (as we found above).

Table 14.1 Six sentence types in the time course studies on Some, the rates of correct (justifiable) responses produced and their concomitant reaction times (from Bott and Noveck, Experiment 3)

Sentence	Example	Justifiable Responses (percentage)	Reaction Time
T1	Some elephants are mammals	True (41%)	2617 (Logical)
		False (59%)	3360 (Pragmatic)
T2	Some mammals are elephants	True (89%)	2644
T3	Some elephants are insects	False (93%)	2610
T4	All elephants are mammals	True (87%)	2875
T5	All mammals are elephants	False (97%)	2558
T6	All elephants are insects	False (92%)	2340

As can be seen in Table 14.1, the main finding here is that mean reaction times were longer when participants responded pragmatically to the Underinformative sentences than when they responded logically. Furthermore, pragmatic responses to the Underinformative sentences appear to be slower than responses to all of the control sentences, indicating that the scalar inference prompts an evaluation that is

characteristically different from all the other items. Collectively, these two experiments provide further evidence against the default inference view because there is no indication that Underinformative items prompt participants to take more time to arrive at a true response than they do to a false response. All indications point to the opposite being true: A logical response is an initial reaction to Underinformative sentences and it is indistinguishable from responses to control sentences while a pragmatic response to Underinformative items is significantly slower than a logical response as well as to the other items in the task.

Some colleagues wonder whether the Pragmatic responses we record are slow because in general judging some statement to be false takes more time than judging it to be true. We argue against that by pointing out that three of the control sentences also require a "false" response and their reaction times are significantly faster than the Pragmatic response. Consider the sentences we classify as T5, which is exemplified by *All mammals are elephants* (which includes many of the same elements as *Some elephants are mammals*). Such items prompt 97% of participants to respond False correctly and at a speed that is significantly faster than when they respond False to Underinformative sentence.[5] It also cannot be argued that the Pragmatic responses are simply due to error (where participants intend to say "True" but hit the wrong button) because the percentage of participants making Pragmatic responses is of a

[5] We have since conducted a follow-up experiment that addresses the key-press issue directly. While using Rips's paradigm, the response options become "agree" or "disagree"and participants first read a statement such as "Mary says that the following is false" or "Mary says that the following is true." Those who *agree* with Mary when she says "false" while keeping a pragmatic interpretation still take significantly longer to respond than those who *agree* when she says "true" while keeping a logical interpretation.

characteristically different order when compared to those making errors in the control conditions (roughly 60% choose False to the Underinformative item as opposed to 3-13% who make errors across all the control conditions). We argue that these results indicate that the scalar inference is at the root of the extraordinary slowdown in this paradigm. It is drawn specifically in reaction to the underinformative items and prompts an unusually large number of participants to ultimately choose False. Furthermore, it arrives as a secondary process relative to a justifiable logical interpretation; it does not appear to arrive by default.

Although our experiments provide evidence against the idea that scalar inferences become available as part of a default interpretation, they do not necessarily provide evidence that directly supports the Relevance account. Our third experiment, which adopted the general procedure of both Experiment 1 (sentences were presented one word at a time) and Experiment 2 (participants were free to treat *Some* as they wished), was designed to test predictions from Relevance theory concerning the processing of scalar inference. The crucial manipulation was the time available for the response; in one condition participants have a relatively long time to respond (3000 msecs, this is referred to as the Long condition), while in the other they are given a short time to respond (900 msecs, this is referred to as the Short condition). By requiring participants to respond quickly in one condition, we limited the cognitive resources they have at their disposal.

Expectations based on the Relevance account are that participants would be more likely to respond with a quick "True" response when they have less time than when they have more. If one wanted to make predictions based on default interpretations, *some*

should be interpreted to mean *some but not all* more often in the short condition than in the long condition (or at least there should be no difference between the two conditions). The results show that while the rate of correct performance among the control sentences either improves or remains constant with added response time, responses were such that there were significantly more Logical responses to the Underinformative items in the Short condition than in the Long condition: 72% True in the Short condition and 56% True in the Long condition. This trend is in line with predictions made by Relevance Theory.

5.2 Disjunctions

This time-enriched pragmatic effect has been extended to disjunctions. Imagine that an experiment presents a five letter word and participants are required to respond with a *Yes* or *No* to statements such as *There is a T or a B*. As one could imagine, all possible truth-conditions can be introduced, including the one synonymous with the underinformative items above. That is, if the word that preceded the statement above were TABLE, then the statement could provide a Logical interpretation (to say "True" because indeed there is a T; there is a B) or a Pragmatic interpretation (to say "False" because there is both a T and a B). In work carried out in collaboration with Valentina Lanzetti, Lewis Bott & Dan Sperber, one sample reveals that when participants are encouraged to respond within a second (they are told that their responses are not recorded in cases where they take longer than a second), participants' rates of Logical responses are roughly 84%. When they are given an unlimited amount of time to decide, their rates of Logical responses drop to around 55%. As before, it is important to point

out that the unambiguous control problems (in this case, half of all included statements using the connective *and*) yielded rates of correct responses that were very high across all conditions. The evidence indicates that indeed minimal interpretations serve as the basis for quick judgements and that pragmatic responses arrive subsequently.

6. Neuropsychological measures and pragmatic inference making

6.1 Evoked Potentials

In order to get at immediate reactions that might run deeper than reaction times, colleagues and I have also investigated Evoked Response Potentials (Noveck & Posada, 2003). ERP studies typically present specific anomalies in a sentence in order to capture a characteristic pattern that follows (see Coulson, Chapter 9). Kutas and Hillyard (1980) pointed out how semantic anomalies give rise to a central parietal negative-going component that peaks about 400 msec after the appearance of an inappropriate word, like *socks* in (7); this is known as an N400:

(7) John buttered his bread with *socks*.

Kutas and Hillyard discovered that N400's are steeper when a target word (which would be the final word here) is (a) not associated with the prior context or; (b) is just unanticipated. Noveck and Posada (2003) investigated N400's as it presented participants Smith's (1980) task (the one that presents *Some elephants have trunks* as an Underinformative item). It included adults only (19 of them) and the quantifier *Some* was the only one used. The task was expanded to include 25 Underinformative sentences

along with 25 sentences that are Patently True (e.g. *Some houses have bricks*) and 25 that are Patently False (e.g. *Some crows have radios*). Before getting to the ERP data, which focused on the final word of each sentence, it is important to highlight the behavioral data.

The reaction-time data are compatible with those found in Bott and Noveck in that False responses to the Underinformative statements take much longer than (nearly twice as long as) the true responses and in that Patently False items yielded the fastest response times overall. Thus, the relative slowness of the *False* responders to the Underinformative items occurs despite evidence of preparedness for the Patently False items. This finding makes it difficult to argue that Underinformative items, by representing one third of the stimuli, allowed for a rote response among the pragmatic responders. What is clear in this study (and less so in Bott and Noveck) is that participants clearly manifest two sorts of strategies: There are 7 who respond true to the Underinformative items by responding literally and quickly overall and there are 12 who respond false by responding pragmatically and slowly overall. It is also apparent that the two strategies prompt spillover because those who respond true to the Underinformative items are also significantly faster in responding correctly to the two other conditions. The behavioral data thus indicate that those who give a false response to the Underinformative items undertake deeper processing that is, in turn, evident in the responses to the other items in the task. Again, the deeper processing linked to the false responses in the Underinformative condition is consistent with expectations based on Relevance Theory because it assumes that pragmatic inference-making arrives as a result of an effortful process.

The ERP data were instructive because they indicated that the Underinformative items generally led to flat N400's (indicating little semantic integration) and were flatter than *both* the Patently True and Patently False items. This is seen regardless of adopted strategies. For those (7 participants) who generally responded true to the Underinformative items and for those (12 participants) who were generally *pragmatic* in their responses, the N400's were comparable. The fact that the ERP profiles for the pragmatic group of participants in the Underinformative condition remains unremarkable, even as their responses and response times indicate much deliberation, is further evidence that participants' immediate reaction to underinformativeness (and signalled by the final word in each sentence) is benign. This indicates that the scalar inference, which requires more effort and prompts participants to respond false, is part of a late-arriving, effort-demanding decision process (see Heinze, Muente, & Kutas, 1998, for a similar argument with respect to a categorization paradigm).

6.2 fMRI

More comprehensive work using imaging techniques is underway with Vinod Goel. The interest here is to see which brain regions are involved in the true vs. false responses (the logical vs. pragmatic responses) to Underinformative items. One study that is now being carried out includes both *All* and *Some* as quantifiers and a large set of stimuli (similar to those in Bott and Noveck, such as *Some monkeys are mammals*).

These are some of the questions that this fMRI work is ultimately meant to address: Right frontal areas are known to be the centers of reconciliation of conflicting information (see Goel & Dolan, 2003; Noveck, Goel, & Smith, in press) as well as for

"contextual" and figurative meanings (Bookheimer, 2002); will a negative reply to underinformative items prompt activity in an area such as in the right anterior cingulate, which is engaged in tasks that require planning and satisfying goals and subgoals (Koechlin, Basso, Pietrini, Panzer, & Grafman, 1999) or will pragmatic inferences prompt activity uniquely in language areas like Broca's area? Will a false response to *Some monkeys are mammals* prompt activity that is different from what is found for a patently false item like *Some monkeys are fish*? If so, one might be able to localize brain regions that concern themselves with pragmatic inference-making. Is neural activity prompted by pragmatic inferencing with scalars similar to other kinds of pragmatics-induced neural activity (e.g. to metaphor and other pragmatic enrichments)? This is an area of research that is in infancy, but will surely be tackled with the expansion of experimental pragmatics.

7. Conclusions

The developmental data show that the production of scalar inference is possible for children (consider seven year olds who are largely "pragmatic" with simple tasks in Pouscoulous et al.) but effortful (now consider seven-to-ten year olds on similar but slightly more difficult tasks in Pouscoulous et al. and in Noveck, 2001 who are "logical"). If a scalar were indeed a default inference linked to a lexical item, adding a little complexity to a task ought not to block it so readily. It is hard to argue that children are more likely to cancel scalar inferences than adults because that implies that children are applying more effort than adults. The evidence shows that young children whose

responses do not reveal the drawing of the scalar inference have simply not considered it.

The adult time-course data tell a similar story. Adults are generally equivocal about statements such as *Some elephants are mammals*, *some elephants have trunks* or that *there is a T or a B* in the word TABLE. It becomes relevant to know which interpretation is the first, competent and measurable one. A default treatment of scalars would predict that it is the upper-bounded *some* (*Some but not all*) and an exclusive interpretation of *or* (exclusive disjunction), but this is not what we find. We consistently find that -- in terms of time-course -- the earliest treatment of such expression is the minimal meaning. The same appears to extend to cases concerning the pragmatic enrichment of *and*. [6] This fits with Relevance Theory's account.[7]

7.1 A final word about the role for experimentation in linguistic-pragmatics

---

[6] One could argue that it is the production of negative information (*not both, not all* etc.) that is the very cause of the slowdown in a scalar inference. However, note how the pragmatic enrichment of *and* also prompts a slowdown (when compared to a "logical" interpretation) yet it cannot be attributed to negative information. The enrichment in these studies concerns adding a sequential marker such as *then*. This is an indication that it is the pragmatic step itself that slows participants down and not necessarily its output.

[7] The grammatical phenomena that Chierchia highlights in Chapter 13 and that led to the investigation in Noveck, Chierchia, Chevaux, Guelminger & Sylvestre (2002) is compatible with an account that a) assumes that there is one minimal meaning of scalar terms that is selectively enriched and that b) denies a role for defaults concerning scalar inferences. Relevance Theory can account for Chierchia's grammatical cases when one views the production of scalar inference as a relatively effortful step that makes an utterance more informative. Informativeness also explains why the lack of scalar inference is more common in the antecedent of conditionals, as in *If P or Q then R*; a truth table analysis shows that there are fewer true cases when the disjunction is inclusive.

As these data reveal, experimentation provides a start for a, potentially long, process that can eventually adjudicate between two theoretical approaches. This is not the only use for experimentation in pragmatics. A subtext of this work is that experimentation can have two other functions in linguistic pragmatics. One is that linguistic intuitions can be usefully complemented with other kinds of evidence. This much might seem obvious enough. The other is that experimentation can also help identify linguistic-pragmatic categories. For instance, the fact that the temporal aspects of the conjunction are attained over the course of development much like scalar implicatures might indicate that these two belong to the same linguistic-pragmatic category. Whether one wants to strengthen or contradict linguistic intuitions, to establish typologies, or devise tests to compare two theories, experimentation has much to offer linguistic pragmaticists and, as the present work shows, linguistic pragmatics has much to offer us experimentalists.

References

Bach, K. (1994) Conversational impliciture. *Mind and Language* 9: 124-62.

Bookheimer, S. (2002). Functional MRI of language: new approaches to understanding the cortical organization of semantic processing. *Annual Review of Neuroscience*, *25*, 151-88.

Bott, L., & Noveck, I. A. (2003). Some utterances are underinformative: The time course of a scalar inference. *Under revision*.

Braine, M. D., & Rumain, B. (1981). Development of comprehension of "or": Evidence for a sequence of competencies. *Journal of Experimental Child Psychology, 31*(1), 46-70.

Carston, R. (1996). *Enrichment and Loosening: Complementary Processes in Deriving the Proposition Expressed?*

Carston, R. (2002). *Thoughts and Utterances*. Oxford: Blackwell.

Evans, J. S. B., & Newstead, S. E. (1980). A study of disjunctive reasoning. *Psychological Research, 41*(4), 373-388.

Goel, V., & Dolan, R. J. (2003). Explaining modulation of reasoning by belief. *Cognition, 87*(1), B11-B22.

Grice, H. P. (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.

Heinze, H. J., Muente, T. F., & Kutas, M. (1998). Context effects in a category verification task as assessed by event-related brain potential (ERP) measures. *Biological Psychology, 47*(2), 121-135.

Horn, L. R. (1973). *Greek Grice: A brief survey of proto-conversational rules in the History of Logic.* Paper presented at the Proceedings of the Ninth Regional Meeting of the Chicago Linguistic Society, Chicago.

Koechlin, E., Basso, G., Pietrini, P., Panzer, S., & Grafman, J. (1999). The role of the anterior prefrontal cortex in human cognition. *Nature, 399*, 148-151.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*, 203-205.

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

Levinson, S. (2000). *Presumptive meanings: The theory of generalized conversational implicature.* Cambridge, MA: MIT Press.

Meroni, L., Gualmini, A., & Crain, S. (2001). *Conversational implicatures and computational complexity in child language.* Paper presented at the LSA, Washington, D.C.

Noveck, I., & Chevaux, F. (2001). *The Pragmatic Development of and.* Paper presented at the BUCLD, Boston.

Noveck, I., Goel, V., & Smith, K. W. (in press). The neural basis of conditional reasoning with arbitrary content. *Cortex*.

Noveck, I., & Posada, A. (2003). Characterizing the time course of an implicature: An evoked potentials study. *Brain and Language, 85*, 203-210.

Noveck, I. A. (2001). When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition, 78*(2), 165-188.

Noveck, I. A., Chierchia, G., Chevaux, F., Guelminger, R., & Sylvestre, E. (2002). Linguistic-pragmatic factors in interpreting disjunctions. *Thinking and Reasoning, 8*(4), 297-326.

Papafragou, A., & Musolino, J. (2003). Scalar implicatures: experiments at the semantics-pragmatics interface. *Cognition, 86*(3), 253-282.

Paris, S. G. (1973). Comprehension of language connectives and propositional logical relationships. *Journal of Experimental Child Psychology, 16(2)*, 278-291.

Rips, L. J. (1975). Quantification and semantic memory. *Cognitive Psychology, 7*(3), 307-340.

Smith, C. L. (1980). Quantifiers and question answering in young children. *Journal of Experimental Child Psychology, 30(2)*, 191-205.

Sperber, D., & Wilson, D. (1986/1995). *Relevance: Communication and Cognition*. Oxford: Basil Blackwell.

Sternberg, R. J. (1979). Developmental patterns in the encoding and combination of logical connectives. *Journal of Experimental Child Psychology, 28*(3), 469-498.