Embedded implicatures?!?

Bart Geurts

Nausicaa Pouscoulous

Abstract

Over the last decade, various proposals have been made for supplanting the classical theory of scalar implicature with a conventionalist (i.e. lexicalist or syntax-based) treatment. In contradistinction to the classical view, conventionalist theories predict that scalar inferences occur systematically and freely in embedded positions. We present experimental evidence that disproves this prediction, arguing along the way that there are rather good reasons for suspecting that introspection is an unreliable tool for gathering data on pragmatic inferences.

Introduction

It is widely agreed that the interpretation of scalar expressions like "some", "or", "good", and so on, involves two rather different ingredients—though recently it has been argued that they aren't that different, either. To explain, consider the following sentence:

(1)   Anna ate some of the cookies.

According to the standard view, the *literal meaning* of (1) is that Anna ate *at least* some of the cookies, which in itself does not rule out the possibility that she ate them all. However, in many cases the sentence will be construed as excluding this possibility, and if it does, there is a *scalar inference* (SI) to

the effect that Anna didn't eat all the cookies. The current debate over the interpretation of scalar expressions is primarily concerned with the status of such SIs.

On the classical Gricean view, SIs are conversational implicatures, which are derived on the assumption that the speaker is trying to be cooperative (Horn 1972, 1989). In a nutshell, the idea is that if the speaker believed that Anna ate all the cookies, (1) would be an uncooperative thing for him to say, and therefore he probably believes that she didn't.[1]

It has been known since Cohen (1971) that the Gricean account of conversational implicature is faced with the awkward fact that, occasionally, conversational implicatures would appear to arise within the scope of expressions like "believe", for example:[2]

(2)   Bob believes that Anna ate some of the cookies.

It seems possible to interpret (2) as conveying that, according to Bob, Anna ate some but not all of the cookies. However, this cannot be explained without further ado as a conversational implicature. To see why not, suppose we try to reason as we did in the case of (1): we observe that, if the speaker thought that Bob believes that Anna ate all the cookies, he would have said so, and since he didn't, we infer that:

(3)   It is not the case that Bob believes that Anna ate all the cookies.

The problem is that this is weaker than the inference that needs to be accounted for. Whereas (3) is true if Bob doesn't have an opinion one way or the other as to whether Anna ate all the cookies, the inference we're after is that Bob believes that Anna didn't eat all the cookies. More generally, in its pristine form, the standard pragmatic account cannot explain SIs arising within the scope of attitude verbs, conditionals, and so on.

---

1. We should note that, in this paper, the opposition between neo-Gricean and relevance theoretic accounts of SIs is not at issue. As such, relevance theory doesn't rule out the possibility that SIs are conversational implicatures, and indeed Noveck and Sperber (2007) maintain that at least some SIs are derived this way. They also say that most SIs arise in different ways, but (and this is our main point) the relevance theoretic framework they adopt doesn't necessitate this position. Be all this as it may, for the purposes of this paper, relevance theory may be regarded as a Gricean framework.
2. Cohen didn't discuss this particular type of example. As far as we are aware, the problem with belief reports was first noted by Landman (2000).

As early as 1971, Cohen used observations like these to argue against the Gricean theory of conversational implicature, but it is only in the last decade that his line of argument has begun to receive broader support.[3] According to what we will call the "conventionalist" view, SIs aren't so much inferences as special information packets which are associated with scalar expressions by linguistic convention. Conventionalism comes in two main varieties: lexicalist and syntax-based. Simplifying somewhat, on a lexicalist analysis "some" is ambiguous between "some and possibly all" and "some but not all". If this is so, (2) is ambiguous, as well, one of its readings being "Bob believes that Anna ate some but not all of the cookies". Thus the problem with (2) is solved by fiat.

Syntax-based members of the conventionalist family secure the same result by rather more formidable means. They stipulate that syntax is equipped with a silent operator, which we will call "**So**", whose meaning is roughly that of overt "only". Supposing that the **So**-operator is adjoinable to any S-node, (2) is predicted to have at least three readings: besides a **So**-free parse, we have:[4]

(4) a. Bob believes that **So**[Anna ate some of the cookies]
    b. **So**[Bob believes that Anna ate some of the cookies]

---

3. The key references are: Landman (1998, 2000), Levinson (2000), Recanati (2003), Chierchia (2004, 2006), Fox (2007), and Chierchia et al. (2008). Although Levinson is usually counted with the "neo-Griceans", his position is closer to that of Landman and Chierchia (2004) than it is to Horn's, for example. It is mainly to circumvent classificatory issues like this that we will avoid using the epithet "neo-Gricean".

4. In fact, unless **So**-insertion is prevented from applying recursively, every sentence has infinitely many parses. For example, for "Anna ate some of the cookies" we get:

   [Anna ate some of the cookies]
   So[Anna ate some of the cookies]
   SoSo[Anna ate some of the cookies]
   SoSoSo[Anna ate some of the cookies] ... and so forth

   So-reiteration tends to be truth-conditionally inert, but it doesn't have to be. The recursive deployment of **So** is the key ingredient in Fox's (2007) analysis of the free choice readings of "You may have an apple or a pear", which according to Fox require parses like the following:

   SoSo[You may [have an apple] or [have a pear]]
   SoSo[You may **So**[have an apple] or **So**[have a pear]]

So is interpreted as a function which takes as its arguments a proposition ϕ and a set of alternative propositions A and returns the conjunction of ϕ and the negation of every proposition in A that is stronger than ϕ (there are other ways of defining the meaning of **So**, but this one will do for the time being). Assuming, to keep things simple, that the alternatives for "Anna ate some of the cookies" are just "Anna ate some of the cookies" and "Anna ate all of the cookies", (4a) is interpreted as:

(5) Bob believes that Anna ate some of the cookies $\wedge$
     Bob believes that $\neg$[Anna ate all of the cookies]

Or, more succinctly, "Bob believes that Anna ate some but not all of the cookies"; which is the problematic reading of (2). If, alternatively, this sentence is parsed as in (4b), and we assume again that there are just two alternatives, then its interpretation is:

(6) Bob believes that Anna ate some of the cookies $\wedge$
     $\neg$[Bob believes that Anna ate all of the cookies]

This is the reading that the standard Gricean analysis accounts for, as well; it is out of reach for lexicalist theories, which can only account for the reading in (5).

If we go shopping for a conventionalist theory of SIs, several choices present themselves. First, we have to decide between the lexicalist and the syntax-based approach. The former is more economical than the latter, but has less predictive power, and it is for this reason, presumably, that recent theories tend to go syntactic. Second, as we mentioned already, if we adopt a syntax-based theory, the precise definition of **So** becomes a matter of concern, but for the time being we will put this issue on one side. For the purpose of this paper, the third choice point is critical. As described so far, conventionalist theories predict merely that sentences containing scalar expressions are systematically ambiguous, with the syntax-based variants yielding more readings than the lexicalist ones. It is generally felt, though, that a theory of SIs should take a stance on what kind of interpretation is most likely to occur. Here, again, there are choices to be made. If we have decided to embrace a lexicalist theory, we could stipulate that the SI-reading of a scalar expression is triggered *by default*.[5] On this view, it should be the case that

---

5. Levinson (2000) adopts an even stronger position, claiming as he does that SIs are not just default inferences but also fast and automatic. This strong variety of defaultism has

in the normal course of events "some" is construed as "some but not all", and the preferred interpretation of (2) is (5). Alternatively, we could stipulate that SI-readings are preferred under such-and-such circumstances, e.g., when a scalar expression doesn't occur in a downward-entailing environment (Landman 1998, 2000, Chierchia 2004).

If, on the other hand, we adopt a syntax-based version of conventionalism, another tack will have to be taken. Chierchia (2006) suggests that whenever **So**-insertion leaves us with several parses to choose from, the one with the strongest meaning is to be preferred; various ways of fleshing out this idea are canvassed by Chierchia et al. (2008). In the case of (2) this theory, too, predicts that (5) is the preferred reading. More generally, conventionalist theories tend to agree that, at least in upward-entailing environments, there is a preference for interpreting embedded "some", in effect, as "some but not all", even if they disagree about how the construal and the preference come about. (This generalisation is a bit brazen, but good enough for our purposes.[6]) Belief sentences are a case in point, as we have seen. Another example are sentences in which a scalar expression is embedded under a strong quantifier: most conventionalist theories agree that (7a) is preferably construed as (7b):

(7) a. {All/Most} students read some of Chierchia's papers.
     b. {All/Most} students read some but not all of Chierchia's papers.

It will be convenient to have a name for this type of construal, so let's say that it involves a "local SI". This term is meant to be purely descriptive: we will say that "some" gives rise to a local SI whenever it is construed *as if* it

---

been the target of several experimental studies which, in our opinion, have shown it to be false (see, in particular, Bott and Noveck 2004 and Breheny et al. 2006, whose findings are corroborated by acquisition data presented by Pouscoulous et al. 2007). However, this debate is not entirely germane to the purposes of this paper, because it doesn't affect weaker versions of defaultism, like Landman's (1998) or Chierchia's (2004), who merely hold that scalar expressions give rise to SIs *ceteris paribus*, without concluding from this that SIs are reflex-like.

6. Here is an example on which the two theories deviate:

(i)   You may have some of the cookies.

On a lexicalist account, the default interpretation of (i) is, "You are permitted to have some but not all of the cookies". For a **So**-theory the strongest reading is, "You may have some of the cookies but you may not have all the cookies".

meant "some but not all". It is not to imply that the lexicalist version of conventionalism is correct; in fact, it is not even to imply that the etiology of local SIs is local in any sense of the word.

While all conventionalist theories predict local SIs and most designate them as preferred in a wide range of cases, the Gricean account of SIs does neither, at least not in its vanilla version. In recent years, it has been argued, however, that on the basis of assumptions that are justified on independent grounds, the Gricean approach can account for *some* instances of local SI (see Geurts 2009 for a conspectus). Crucially, however, such explanations are strictly piecemeal: unlike its conventionalist competitors the Gricean framework does not provide a general-purpose mechanism for generating local SIs across the board. For example, whereas several proposals have been made as to how SIs can have (or rather, can seem to have) narrow scope with respect to "believe", these proposals do not extend to other atttitude verbs or any other form embedding, like strong quantifiers. We will briefly revisit this topic in Section 8. Until then, the main thing to bear in mind is that, whereas for conventionalist theories local SIs are systematic, if not the rule, for Gricean theories they are exceptional.

The family of conventionalist theories surveyed in the foregoing represent what we call "mainstream conventionalism", for the simple reason that most conventionalists adhere to this view (a possible exception is Fox 2007). The primary goal of this paper is to argue, on the basis of experimental evidence, that mainstream conventionalism is wrong. One way of saving conventionalism (if one is so inclined) is by leaving the mainstream and contenting oneself with the position that sentences containing scalar expressions are systematically ambiguous, without claiming a preference for any particular type of reading. We believe that even such a minimal version of conventionalism can be refuted, and will back up this claim with experimental evidence (Section 7).

We should like to emphasise that the main objective of this paper is to show that conventionalism is wrong, not to prove that the Gricean approach is right. Still, we believe that our data are more in line with the Gricean view, and in our concluding section we will explain why.

1   Experiments 1a and 1b

The aim of our first two experiments was to obtain an initial impression of how SIs behave under embedding. Since the two experiments were based on the same design and were conducted separately for practical reasons only, we present them together. Both experiments compared the interpretation of "some" occurring in simple sentences with occurrences of the same word embedded in the scope of various expressions. In Experiment 1a, "some" was embedded under "think", deontic "must", and the universal quantifier "all"; in Experiment 1b the complex conditions featured "think" and "want".

1.1   Participants

The participants in Experiment 1a were 30 French-speaking first-year students at the Ecole Nationale des Arts Décoratifs in Paris. Experiment 1b drew 31 participants from the same population. There was no overlap between the two groups.

1.2   Method and procedure

In both experiments, participants were presented with questionnaires displaying on each page a task like the one shown in Figure 1. Written instructions explained the task with the help of an inference that didn't involve scalar

---

Emilie says:

"Betty thinks that Fred heard some of the Verdi operas."

Would you infer from this that Betty thinks that Fred didn't hear all the Verdi operas?

□ yes   □ no

Figure 1: Trial used in Experiments 1a-b.

expressions, and participants were encouraged to be accurate but not dwell too long on any given item (the same formula was used in all the studies reported in this paper).

We refer to Table 1 for a sample of the tasks that were used. In both experiments, participants saw one control item in which "some" occurred in a non-embedded position. In Experiment 1a there were three conditions that had "some" occurring in an embedded position, i.e. in the scope of "think", "all", and deontic "must". In Experiment 1b, there were two complex conditions: one with "think" and one with "want".

Filler items were used to conceal the purpose of the experiment; the number of filler items was 16 in Experiment 1a and 9 in Experiment 1b. Filler items had the same general format as critical items, but didn't contain scalar expressions. Whereas the critical and filler sentences were always of the same form, content words were different for each participant; hence, within either experiment, every critical item was used only once. Furthermore, each participant saw the materials in a different, pseudo-random order, with at least two filler items separating any pair of critical items.

	target sentence	candidate inference
∅	Fred heard some of the Verdi operas.	He didn't hear all of them.
all	All students heard some of the Verdi operas.	None of the students heard them all.
must	Fred has to hear some of the Verdi operas.	He isn't allowed to hear all of them.
think	Betty thinks Fred heard some of the Verdi operas.	She thinks he didn't hear all of them.
want	Betty wants Fred to hear some of the Verdi operas.	She wants him not to hear all of them.

Table 1: Sample sentences used in Experiments 1a-b.

The sentences in the right-most column of Table 1 are inferences that, on a mainstream conventionalist account, follow immediately from the preferred interpretations of the target sentences. For example, if "some" is interpreted with a local SI, "All students heard some of the Verdi's operas" means that all of them heard some but not all of Verdi's operas, which is to say that none of the students heard them all. The same, mutatis mutandis, for the other sentences.

1.3 Results

While in both experiments SIs were endorsed over 90% of the time in the simple ∅-conditions, in the complex conditions these rates were considerably lower; see Table 2. Cochran's Q-test confirmed that the critical conditions differed from each other (Experiment 1a: Q = 49.750, n = 30, df = 3, p < .001; Experiment 1b: Q = 21.68, n = 31, df = 2, p < .001). Further pairwise comparisons revealed that in both experiments the ∅-condition differed from

	∅	*all*	*must*	*think*	*want*
Experiment 1a	.93	.27	.03	.50	–
Experiment 1b	.94	–	–	.65	.32

Table 2: Rates of positive responses observed in Experiments 1a-b.

nearly all complex conditions (McNemar's p ≤ .001 for all comparisons, except for ∅ vs. "think" in Experiment 1b, where p = .012). Clearly, the rates at which SIs were derived in the complex conditions were much lower than in the ∅-conditions. Furthermore, in both experiments all complex conditions differed significantly from each other (Experiment 1a: "all" vs. "think": p = .039; "all" vs. "must": p = .039; "think" vs. "must": p = .001; Experiment 1b: "think" vs. "want": p = .031).

1.4 Discussion

Prima facie, these results are inconsistent with mainstream conventionalism. First, and most importantly, whereas our data are consistent with the view that SIs are the default in simple sentences, the complex conditions yielded local SIs at a much reduced mean rate of 35%. Secondly, we observed fluctuations between complex conditions: local SIs were relatively frequent with "think" (57.5% across the two experiments), practically non-existent with "must" (3%), and relatively rare with "all" (27%) and "want" (32%). These data indicate that mainstream conventionalism is on the wrong track.

There are, as far as we can see, two ways in which the conventionalist might try to escape this conclusion. One would be to argue that, in the complex conditions, local SIs were suppressed because they yielded implausible interpretations. Another way would be to point to the undeniable fact that complex sentences are more complex than simple ones, and attribute the low

rates of local SIs to increased processing demands. In the following we will argue that neither line of defense cuts any ice.

We believe that the implausibility argument fails, for the following three reasons:

Objection #1: It is clear that that the appeal to implausibility will not work for embedding under "all" or "thinks":

(8)   All students heard some of the Beethoven symphonies.
  a.   All students heard some but not all of the Beethoven symphonies.
  b.   All students heard some and maybe all of the Beethoven symphonies.

(9)   Betty thinks that Fred heard some of the Beethoven symphonies.
  a.   Betty thinks that Fred heard some but not all of the Beethoven symphonies.
  b.   Betty thinks that Fred heard some and maybe all of the Beethoven symphonies.

In (8) and (9) the (a) and (b) sentences paraphrase the construals with and without local SIs, respectively. In Experiment 1a, sentences like (8) yielded local SIs only 27% of the time. Hence, in order for the plausibility argument to work, (8a) should be markedly less plausible than (8b). According to our intuitions, this is simply false. The same holds for (9), though this case is admittedly somewhat less clear cut. Across the two experiments, sentences like (9) yielded local SIs at a rate of 57.5%. Hence, in order for the plausibility argument to work, about half of the time participants must have felt that (9a) is sufficiently implausible to override the preference for local SIs. Unless this preference is so weak that it can hardly be called that, this prediction doesn't seem to be right, either.

Actually, the argument from implausibility will have to be more sophisticated than the last paragraph gives it credit for. If we want to save the predictions made by mainstream conventionalism, it will not be enough to argue that the (a)-construals are significantly less plausible than the (b)-construals. Rather, what needs to be shown is that the likelihood that the speaker believes the former is significantly smaller than the likelihood that he merely believes the latter. But again, we see absolutely no reason for accepting that this is the case.

Objection #2: Against the first objection, it might be objected that even if the argument from implausibility doesn't go through for "all" and "think", it is just right for the "must" and "want" cases:

(10) Fred has to hear some of the Beethoven symphonies.
    a. Fred has to hear some but not all of the Beethoven symphonies.
    b. Fred has to hear some and maybe all of the Beethoven symphonies.

(11) Betty wants Fred to hear some of the Beethoven symphonies.
    a. Betty wants Fred to hear some but not all of the Beethoven symphonies.
    b. Betty wants Fred to hear some and maybe all of the Beethoven symphonies.

(Note that in the paraphrases "not all" and "maybe all" are meant to have narrow scope.) Surely, the (a) readings are less plausible than the (b) readings, so in these cases, at least, it is natural to suppose that the preference for local SIs is overridden, and mainstream conventionalism is off the hook.

This argument is a specious one. The argument presupposes that preferred SIs will quietly withdraw whenever they happen to be implausible, and this presupposition is doubtful, since it is contradicted by the behaviour of bona fide SIs. Let us consider a number of examples to back up this claim.

(12) Some of the liberal MPs voted against the bill.

In some democracies, parliamentary parties tend to vote *en bloc*. In the Netherlands, for instance, it would be quite unlikely that some but not all liberal MPs voted against a bill. Nevertheless, as far as we can see, this wouldn't diminish the likelihood that (12) was interpreted as implicating that the liberal MPs were divided.

(13) In order to prevent the rinderpest from spreading through his herd, some of Jones's cows were vaccinated.

Since rinderpest is a highly contagious disease, it would be decidedly odd if only some of Jones's cows were vaccinated, yet that is how we would understand (13).

(14) Anna threw all her marbles in the swimming pool. Some of them sank to the bottom.

No doubt, it would be *very* odd if some of Anna's marbles failed to sink, yet according to our intuitions that is precisely what (14) conveys.

As observed by Geurts (1999), it appears that genuine scalar implicatures are not so easy to cancel, at all. It might be thought that in the foregoing examples, SIs are less cancellable because they are triggered in unembedded positions (though we are at a loss to see why that should play a role), but this is not the case. To explain why not, let us start with an uncontroversial example:

(15)   Harry wants to correct some of the figures.

This may, and perhaps normally will, be understood as conveying that it is not the case that Harry wants to correct all the figures (with wide scope for the negation). This observation is undisputed, though we note in passing that it is problematic for lexicalist variants of conventionalism. But now consider the following:

(16)   Harry wants some of the figures to be correct.

Surely, it would be remarkable, to say the least, if it wasn't the case that Harry wanted all the figures to be correct. Still, that's how we would interpret (16), and the following example is perhaps even clearer in this respect:

(17)   Harry wants some of his grandchildren to be happy.

To sum up: the implausibility argument is based on the supposition that the preference for SIs is so weak that it will be overridden whenever the preferred reading is even mildly improbable. The foregoing observations suggest rather forcefully that this supposition cannot be justified on independent grounds.

Objection #3:   One of our findings was that local SIs were relatively frequent with "think" (57.5% on average), practically non-existent with "must" (3%), and relatively rare with "all" (27%) and "want" (32%). If the argument from implausibility is correct, people's plausibility judgements should mirror these differences. There are many different ways of securing this correlation, of course, but in any event it would seem to require that (18a) is significantly more plausible than (18b), which in its turn should be significantly more plausible than (18c):

(18)   a. Betty thinks that Fred read some but not all of the Harry Potter

books.
- b. All the students read some but not all of the Harry Potter books.
- c. Fred has to read some but not all of the Harry Potter books.

Although we cannot rule out that experimental research might confirm these predictions, we consider it quite unlikely.

The upshot of the foregoing discussion is that the implausibility argument, though tempting at first, turns out to be dubious on closer examination. There is, however, another way of bringing our data in line with mainstream conventionalism. The complex conditions in Experiments 1a-b were obviously more complex than the $\varnothing$-conditions, and it might be suggested that this is why we observed such low rates of local SIs. There are at least two problems with this line of argument. One is that it doesn't account for the fact that complex conditions curtail SI rates to varying degrees, unless it can be assumed that "must" sentences are more complex than "all" sentences, that "want" sentences are more complex than "think" sentences, and so on. We see no independent grounds for believing that this is so.

Secondly, although complex sentences are more complex than $\varnothing$-sentences, we doubt that the difference is sufficiently large to account for the steep drop in SI rates. This doubt was confirmed by the following informal experiment. In a classroom setting, we asked 31 Dutch philosophy students to decide if the following inference is valid:

(19)   Mary has to put some but not all of the stamps in a blue envelope.
       SO: She may not put all the stamps in the blue envelope.

27 (or 89%) of our students said that the inference was valid, thus confirming our expectation that the *logical* form of the conventionalist inferences in our experiment wasn't particularly difficult. Against this argument it may be objected that the reasoning in (19) is made easier by the fact that the "not all" constraint is made explicit, whereas in our experiment it was implicit. However, this objection has little force in view of the well-established fact that reasoning with implicit premisses in usually easier than when the premisses are made explicit (see, e.g., Geurts and van der Slik 2005). Furthermore, it should be noted that the *grammatical* form of the premiss in (19) is more complex than that of its counterpart in Experiment 1, which should offset, to some degree at least, such advantages as might be gained from increased explicitness.

Having argued that the results of Experiments 1a-b are at odds with mainstream conventionalism, we ask our readers to brace themselves for an unexpected turn of events: we are about to argue that even the modest rates of local SIs observed in the foregoing experiments are, to some degree at least, experimental artefacts. If this argument is correct, our case against conventionalism becomes even stronger, but it also raises some hairy methodological issues, which we will not shy away from.

2   Worries about the paradigm

As is common in linguistics and philosophy, the literature on SIs is based almost entirely on introspective evidence.[7] We have argued elsewhere that the introspective method is biased when applied to SIs (Pouscoulous 2006, Geurts 2009), and in this section we will expand on these arguments. It will be clear that the inference paradigm adopted in Experiments 1a-b comes down to collecting introspective judgements on SIs from a population of naive native speakers, and there are three reasons for suspecting that this paradigm might yield exaggerated levels of SIs.

Worry #1:   Consider the following argument:

(20)   If Jack is happy, Jill is happy.
       so: If Jill is not happy, Jack is not happy.

When asked, many people would say that this is a valid argument. But surely this does not entail that as many people will spontaneously derive the same conclusion when they learn that Jill is happy if Jack is. In general, the rate at which people will spontaneously draw a conclusion ϕ from a set of premisses A will be lower than the rate at which people are prepared to endorse the corresponding argument "A, therefore ϕ" (see, e.g., Evans et al. 1993).

The implications for the inference paradigm of Experiments 1a-b will be obvious: if we ask people whether the inference in (21) is correct, and they say "yes" in 32% of the cases, as they did in Experiment 1b, then it is a

---

7.   Almost, but not quite: as we mentioned in footnote 5, there have been quite a few experimental studies addressing the question whether SIs are strong defaults (i.e. reflex-like). But as we said before, this experimental work has no direct bearing on the issues we are concerned with here.

moral certainty that the rate at which they would spontaneously draw this inference is lower than 32%.

(21)   Betty wants Fred to hear some of the Verdi operas.
       ⤳ She wants him not to hear all of the Verdi operas.

Worry #2:   Suppose we ask ourselves whether (22b) follows from (22a):

(22)   a. Fred has heard some of the Verdi operas.
       b. Fred hasn't heard all the Verdi operas.

The very question whether (22b) might be implied changes the context in which (22a) is interpreted: the question makes it relevant to decide whether or not the speaker believes that (22b) is true. Hence, while we might not even consider (22b) when (22a) is presented in a more neutral setting, the inferential paradigm is bound to enhance the likelihood that the SI is endorsed.

Worry #3:   Consider the following argument schemes:

(23) | All A are B | All A are B | All B are A | All B are A |
       | --- | --- | --- | --- |
       | All B are C | All C are B | All B are C | All C are B |
       | All A are C | All A are C | All A are C | All A are C |

The left-most syllogism is valid, and many experiments confirm that people are very good at recognising that it is: according to Chater and Oaksford's (1999) meta-analysis it was accepted 90% of the time across a range of independent experiments. The other three arguments in (23) are not valid, yet they are accepted 63% of the time on average. What causes this high error rate? The answer is fairly clear: it is because all the arguments in (23) look alike. This is what we see throughout the space of syllogistic arguments: valid arguments tend to be surrounded by invalid but superficially similar arguments causing high error rates.

Our third worry is that the same might hold for local SIs. There is a consensus that (22a) may give rise to the inference that (22b) is true, and that this type of inference is pragmatically valid. Our worry is that, at least some of the time, people who endorse the inference in (21) may do so because it is superficially similar to the uncontested inference in (22), since prima facie it seems to involve the same form of substitution. If people can

15

commit errors in logical reasoning, they can also commit errors in pragmatic reasoning, and this would be a plausible candidate.

We have given three reasons for being wary of the inference paradigm, and all our worries point in the same direction: it is possible that the inference paradigm yields too many SIs. However, it remains to be seen whether this is a real problem, for even if our worries are justified it may be that the bias caused by inference paradigm is so weak that it can be safely ignored. The next two experiments address this issue.

3   Experiment 2

In this experiment we compared the inference task with a verification task, which has been widely used in experimental research on implicatures (e.g., Noveck 2001, Bott and Noveck 2004, Pouscoulous et al. 2007). As far as we can tell, the verification paradigm does not raise the kind of issues that the inference paradigm gives rise to, and is therefore a more reliable instrument for gauging the rates at which people derive SIs.

Experiment 2 was a small-scale study whose sole purpose was to establish whether or not the inference paradigm yields higher rates of SIs in simple sentences than the verification paradigm. We will return to complex sentences in Experiment 3.

3.1   Method and procedure

29 Dutch-speaking students at the University of Nijmegen were presented with the same sentence in two different tasks: an inference task, like the one used in Experiments 1a-b, and a verification task. The critical sentence was the following:

(24)   Some of the B's are in the box on the left.

In the inference condition, participants had to decide whether this implies that not all the B's are in the box on the left, and were asked to tick "yes" or "no" accordingly, as in Experiments 1a-b. In the verification condition, participants had to decide whether the same sentence correctly describes the following situation:

$$\boxed{\text{B B B A A A}} \qquad \boxed{\text{C C C}}$$

Following Bott and Noveck (2004) and others, we assume that someone who interprets (24) as implicating that not all the B's are in the box on the left will deny that the sentence gives a correct description of the facts. Note that, whereas in the inference task the SI is associated with a positive response, in the verification task it is associated with a negative response; therefore we had to control for the possibility of a positive response bias.

In both conditions, the critical sentence was the same, i.e. (24), and 6 similar items were used as fillers; in the verification tasks filler items were also intended as controls for a potential positive response bias. For this purpose, we used simple statements like, "There are more than two B's in the box on the right", half of which were true, while the other half were false. Thus, each participant saw a total of 14 stimuli. These were divided in two blocks, one with the inference items, the other with the verification items. The order of presentation of the two blocks was counterbalanced across participants, in such a way that 15 participants saw the inference tasks first, while the remaining 14 started with the verification tasks. Within each block, items were presented in a random order, while ensuring that the first three items were fillers.

Participants were given a questionnaire with general instructions printed on the cover, which informed them they would have to solve two different tasks. This page also contained more specific instructions for the task in the first block of trials; the instructions for the second block were printed on a separate page preceding it.

3.2   Results and discussion

We found no order effect for either condition (inference: Fischer's Exact test, p = .45; verification: Fischer's exact test, p = 1). As expected, participants derived SIs more frequently in the inference condition (62%) than in the verification condition (34%) (McNemar's test, n = 29, p < .01). Since participants' performance on the filler items in the verification task was nearly perfect (97% correct), it is unlikely that the low rate of "no" answers for the critical item was caused by a positive response bias. In short, the outcome of this experiment confirms our conjecture: the inference task enhances the rate of SIs, and moreover, it appears that the effect is quite substantial.

We have to concede that, strictly speaking, the outcome of this experiment has no bearing on the issue of embedded SIs. Since the only critical item was a simple sentence, our result merely indicates that the rates of the SIs for control items in Experiments 1a-b may have been too high. In itself, this wouldn't be problematic for conventionalism, though it has to be noted that it wouldn't improve the theory's situation, either. However, as we will see in the next experiment, the discrepancy between the inference and the verification tasks extends to complex sentences, as well.

4 Monotonicity

Everybody agrees that there is no preference for local SIs in downward-entailing contexts:

(25) a. Not all the squares are connected with some of the circles $\not\rightsquigarrow$ Not all the squares are connected with some but not all of the circles.
 b. There isn't more than one square that is connected with some of the circles $\not\rightsquigarrow$ There isn't more than one square that is connected with some but not all of the circles.

All mainstream versions of conventionalism agree that there is a preference for local SIs in upward-entailing contexts like the following:

(26) a. All the squares are connected with some of the circles $\rightsquigarrow$ All the squares are connected with some but not all of the circles.
 b. There is more than one square that is connected with some of the circles $\rightsquigarrow$ There is more than one square that is connected with some but not all of the circles.

Furthermore, mainstream conventionalist theories make the same predictions for non-monotonic contexts:

(27) There are exactly two squares that are connected with some of the circles $\rightsquigarrow$ There are exactly two squares that are connected with some but not all of the circles.

It is not evident that these predictions are correct, and that's putting it mildly. However, despite the fact that mainstream conventionalism predicts that local SIs are the norm in non-downward-entailing (i.e. upward-entailing

and non-monotonic) contexts like (26) and (27), we aren't aware of any attempts at defending this prediction. According to Geurts (2009), the default construal of the sentences in (26) and (27) is without local SI. Van Rooij and Schulz (2004) claim that native speakers' opinions are divided, but argue that local SIs can occur in upward-entailing contexts. Even Chierchia (2004) merely claims that local SIs are possible in examples like (26a,b) and (27), and does not go so far as to argue that it is the preferred option. Nevertheless, the fact remains that Chierchia's theory predicts that it is, and this is what we tested in the next experiment.

5 Experiment 3

In Experiment 3, we tested people's intuitions about the sentences in (25)-(27), pitting a verification task against an inference task, as we did in Experiment 2. The purpose of this experiment was twofold. In the first place, we wanted to test the prediction that the inference paradigm yields higher rates of local SIs than does the verification paradigm, which is what we hypothesised on the basis of the arguments given in Section 2 and the outcome of Experiment 2. A priori, there is no reason why such an effect should differentiate between downward-entailing (DE) and non-downward-entailing (non-DE) contexts, and therefore our prediction was that it should occur across the board. This is to say that, despite the universal consensus that SIs are suppressed in DE environments, we expected to find at least some SIs (or perhaps it is better to say "pseudo-SIs") even in DE environments, in the inference condition.

In the second place, we wanted to test the prediction, made by mainstream conventionalism, that there will be high rates of local SIs in all non-downward-entailing (non-DE) contexts. One of the key tenets of mainstream conventionalism is that we should find high rates of local SIs in these cases. This prediction holds for both types of task, and if it is true that the inference task boosts the rate of SIs, we should find even higher rates in the inference condition, if mainstream conventionalism is right.

5.1 Method and procedure

26 first-year students in the humanities at the University of Nijmegen, all native speakers of Dutch, were presented with two types of tasks, as in Ex-
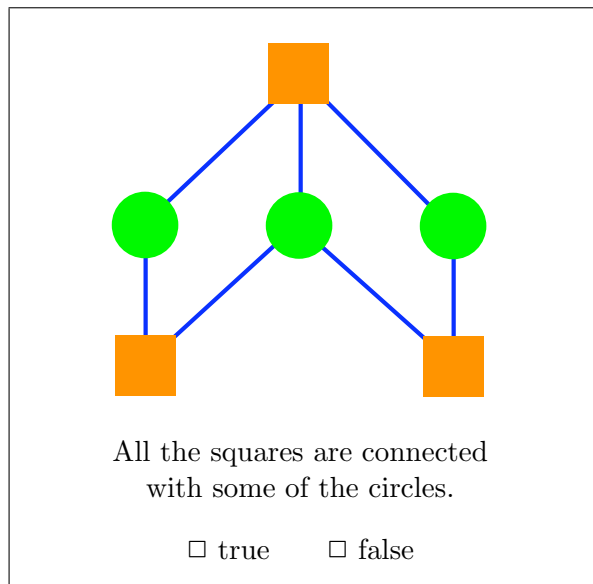
Figure 2: Verification item used in Experiment 3.

periment 2. The critical sentences were the ones in (25)-(27). Samples of verification and inference trials are given in Figures 2 and 3. In the verification condition, each of the critical sentences was paired with a situation in which its classical construal and a local-SI construal yielded conflicting truth values. For example, when interpreted with a local SI, the sentence in Figure 2, i.e. (26a), fails to match the depicted situation, but it is true if "some" isn't strengthened. By the same token, (25a), which is the negation of (26a), is true with and false without a local SI. The same, mutatis mutandis, for the "more than" sentences in (25b) and (26b).

Sentence (27), in which "some" occurs in the scope of non-monotonic "exactly two", is a special case. According to mainstream conventionalism, this sentence is preferably interpreted in such a way that it is true if two squares are connected with some but not all of the circles while one square is connected with all the circles, and false if one square is connected with some but not all of the circles while one square is connected with all the circles. We decided to test both predictions, and therefore included two verification trials with this sentence.

Thus, in the verification task there were 6 critical items altogether. These were mixed with 37 superficially similar items, which were part of two other,

20

```
┌─────────────────────────────────────────────────┐
│                                                 │
│   Betty says:                                   │
│                                                 │
│        "All the squares are connected with      │
│        some of the circles."                    │
│                                                 │
│   Could you infer from this that, according     │
│   to Betty:                                     │
│                                                 │
│        All the squares are connected with       │
│        some but not all of the circles.         │
│                                                 │
│                  ☐ yes      ☐ no                │
│                                                 │
└─────────────────────────────────────────────────┘
```

Figure 3: Inference item used in Experiment 3.

unrelated experiments. So in total there were 43 items, which every participant saw in one of 10 pseudo-random orders.

In the inference task, there were 5 critical trials: one for each of the sentences in (25)-(27). In every instance, the candidate inference was obtained by replacing "some" in the speaker's statement with "some but not all", as in (25)-(27), and changing the word order from verb-second to verb-final; the latter change was necessary because in Dutch that-clauses the word order is verb-final. The critical items were interspersed with 10 superficially similar filler items. Every participant saw the 15 items in one of 10 pseudo-random orders.

The block of verification items was presented first, and it was followed by an unrelated experiment that took about 20 minutes to complete, after which the inference items were presented.

5.2   Results and discussion

Table 3 presents the results of the experiment. In this table, the figures in brackets represent the rates positive responses should tend towards, according to mainstream conventionalism. For example, in the verification condition, the preferred response to the "all" statement should be "false".

In the verification condition, the results for the DE quantifiers ("not all" and "not more than two") were as expected: nearly all participants rejected "Not all the squares are connected with some of the circles" (= (25a)) if all the squares were connected with some of the circles, even if some of the

21

	verification		inference	
all	1	(0)	.46	(1)
more than one	1	(0)	.62	(1)
exactly two	1	(0)	.5	(1)
exactly two	0	(1)		
not all	.04	(0)	.58	(0)
not more than one	.04	(0)	.46	(0)

Table 3: Rates of positive responses observed in Experiment 3. Response trends predicted by mainstream conventionalism are given in brackets. Note that there were two verification conditions with "exactly two" against one inference condition.

squares were connected with all the circles; the same, mutatis mutandis, for sentences with "not more than one". However, since this is predicted by all theories of scalar inference, it hardly counts as evidence in favour of mainstream conventionalism. On the other hand, in the non-DE verification conditions ("all", "more than one", and "exactly two"), the observed response rates were the exact inverse of what the theory predicts. Put otherwise, the verification tasks completely failed to yield the local SIs predicted by mainstream conventionalism.

In the inference condition, participants failed to produce the predicted response pattern, as well, since in this case *all* rates, for DE and non-DE items alike, clustered around chance level, give or take 12%. Pairwise comparisons between inference and verification items showed that the differences were significant throughout (McNemar's test with Bonferroni correction for multiple comparisons: "all": $p < .005$, "not all": $p < .001$, "more than one": $p < .0005$, "not more than one": $p < .05$, "exactly two": $p < .005$ in both cases).

The implications of these results are straightforward. One prediction we set out to test was that local SIs are preferred in non-DE contexts, and this prediction wasn't confirmed by either part of the experiment. In the verification condition, there was no evidence for local SIs at all. In the inference condition, all response rates clustered around 50%, and failed to coalesce into the pattern predicted by mainstream conventionalism: in the DE conditions the rates of positive responses were too high (mean: 45%), while in non-DE contexts they were too low (mean: 51%) to support the prediction that in these contexts local SIs are derived by default. In short,

our data clearly disconfirm mainstream conventionalism.

The second prediction this experiment was designed to test was that, compared to the verification paradigm, the inference paradigm should give rise to elevated levels of local SIs. This prediction was confirmed by mean rates of 1% and 51%, respectively. This finding throws new light on the outcome of Experiments 1a-b: it vindicates our conjecture that the rates of local-SI readings observed in Experiments 1a-b must be attributed at least in part to the inference paradigm. If we correct for this effect, then of course the positive response rates that were relatively low, to begin with, will become lower still, and given the discrepancy between the two paradigms we found in Experiment 3, it is quite possible that the rates previously observed for "all" (27%) and "want" (32%) are entirely due to a paradigm bias. That is to say, it is not unlikely that the "think" condition was the *only* complex condition in Experiments 1a-b whose elevated level of positive responses (57.5%) wasn't merely an artefact.

6 Minimal conventionalism

Thus far, we have been concerned with what we have dubbed "mainstream conventionalism", a label for a spectrum of theories all of which predict that local strenghtening should be preferred in the environments studied in Experiments 1a-b and the non-DE environments of Experiment 3. We presented data showing, arguably, that the predictions of mainstream conventionalism are wrong. What to do? Our poorly hidden agenda is that everybody should return to the pragmatic fold. Another option is to leave the mainstream without giving up on conventionalism: stick to your favourite lexicalist or syntax-based brand of SIs, which will duly generate a batch of interpretations for any sentence that contains scalar expressions, but refuse to make predictions about which construal is the preferred one. Leave it to pragmatics.

To the best of our knowledge, nobody has come forward to advocate such a minimalist take on conventionalism, as yet, and we are tempted to say that this is just as well, since this view is too weak to be taken seriously. However, instead of lambasting it for its lack of ambition, let us consider how minimal conventionalism might be tested. We believe that it is instructive to try doing so, precisely because it is such a weak position: if minimal conventionalism can be shown to be problematic, stronger versions of conventionalism are

liable to share its problems.

Obviously, in its most austere form, minimal conventionalism fails to make any predictions at all: the claim that (28a) may or may not be read as (28b) cannot be falsified:

(28)   a. All the customers shot at some of the salesmen.
       b. All the customers shot at some but not all of the salesmen.

In order for the theory to have at least a modicum of empirical bite, auxiliary assumptions will have to be made. A natural assumption is that native speakers are able to detect the readings distinctive of the conventionalist approach. That is to say, even if speakers don't prefer to hear (28a) as synonymous with (28b), it is reasonable to expect them to be aware that the latter sentence expresses a possible construal of the former. If this auxiliary assumption is made, minimal conventionalism predicts that, in a situation that falsifies (28b), native speakers should claim either that (28a) is false or that it is ambiguous between a true and a false reading. This prediction was tested in our last experiment.

7   Experiment 4

For this experiment, we took the materials of the verification task of Experiment 3, translated them into English, and added a third response option: participants were invited to choose between "true", "false", and "could be either". According to minimal conventionalism, participants should have a strong preference for the last option in the non-DE cases, assuming that they are able to detect the predicted ambiguities. In order to verify that participants can cope with this type of task, we added a number of control sentences which, according to our intuitions, were unambiguously ambiguous.

7.1   Method and procedure

The target items were the same as in the verification condition of Experiment 3, save that (a) they were in English and (b) there was an additional response option, viz. "could be either". A further difference with the previous experiment concerned the status of DE sentences: since the aim of this experiment was just to test conventionalist predictions about non-DE environments, the DE sentences now merely served as controls.
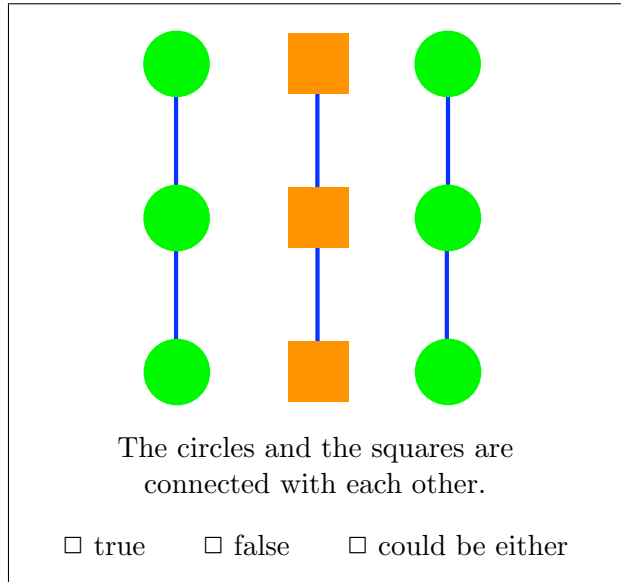
Figure 4: Control item used in Experiment 4.

More important than the DE controls were the following control sentences, which according to our intuitions were clearly ambiguous:

(29) a. The circles and the squares are connected with each other.
      b. The green and the orange figures are connected with each other.
      c. All the figures are orange and green.
      d. There are green circles and squares.
      e. The circles and the squares have the same colour.

A sample control item is shown in Figure 4. In addition to 4 critical non-DE items, 2 DE controls, and 5 ambiguous controls, there were 19 fillers, so the experiment consisted of 30 trials in total, which were presented in 10 different, pseudo-random orders.

The participants in this study were 22 first-year linguistics students at University College London. The experiment was conducted in a classroom setting and instructions were given orally, because we wanted to ensure that the participants understood the notion of ambiguity. This notion was explained with the help of examples like the following:

(30) a. Visiting relatives can be boring.

– It can be boring to visit relatives.

– Relatives who come to visit can be boring.

b. The girl hit the boy with the telescope.

– The girl hit the boy who had the telescope.

– The girl used the telescope for hitting the boy.

For all examples it was shown that, in the right kind of situation, it could be either true or false, i.e. true on one construal and false on another. Participants were told that the aim of the experiment was to assess if and to what extent people can detect ambiguities in English sentences.

7.2 Results and discussion

Table 4 presents the main results of this experiment; the shaded cells in this table contain the rates of positive responses that are inconsistent with minimal conventionalism. Observe, to begin with, that in the DE control conditions ("not all" and "not more than one"), participants overwhelmingly failed to obtain a local-SI reading, which is in line with what we found in Experiment 3 and consistent with just about any theory of scalar inference, including minimal conventionalism. As for the critical non-DE sentences,

	yes	no	both
all	.95	.05	0
more than one	1	0	0
exactly two	.86	.05	.09
exactly two	.09	.77	.14
not all	.09	.86	.05
not more than one	.09	.91	0

Table 4: Response rates for the critical and DE control items in Experiment 4. The percentages in the shaded cells are inconsistent with minimal conventionalism. As in Experiment 3, there were two trials with "exactly two".

whereas ambiguous controls were recognised as such in 70% of the cases (see Table 5), the corresponding rate for the critical non-DE sentences was 6%. This difference was highly significant (Wilcoxon's Exact test: W = 208, n = 20, p < .0001). On a further 5% of the non-DE trials, participants gave a true/false response that was consistent with a local-SI construal. In total, on the non-DE trials, only 9 out of 88 responses (i.e. 10%) were consistent

The circles and the squares are connected with each other	.82
The green and the orange figures are connected with each other	.73
All the figures are orange and green	.59
There are green circles and squares	.77
The circles and the squares have the same colour	.59

Table 5: Response rates for the ambiguous control items in Experiment 4.

with minimal conventionalism.[8] Moreover, all but one of these responses were associated with non-monotonic "exactly two": between them, upward-entailing "all" and "more than one" prompted a local-SI construal on *one trial* only.

To sum up, while they were rather good at picking out undisputedly ambiguous sentences, participants overwhelmingly failed to produce responses consistent with a conventionalist analysis of scalar expressions, either by classifying them as ambiguous or by assigning them a truth value in line with a local-SI construal. In our opinion, this is very strong evidence against the conventionalist approach to SIs.

We set out to test an extremely weak version of conventionalism, and the outcome is clearly negative. Short of recanting their faith altogether, there is only one way out for conventionalists that we can see: they will have to claim that the ambiguities predicted by their theory are actually very hard to fathom. Needless to say, such a claim would have to be motivated, and we don't really see how that might be done. But apart from that it may be remarked that, if this is the conventionalist doctrine of the future, the theory is heading for spectacular turnabout: what started out as a preferred construal ends up being an elusive reading observable under special circumstances only.

8   Meanwhile in the Gricean camp . . .

Although the chief purpose of this paper was to argue that conventionalism is on the wrong track, it may be instructive to consider, if only briefly, how the standard Gricean theory fares with the experimental evidence mustered in this paper. Recall that what motivated conventionalism, in the first place, was what Chierchia et al. (2008) put forward as an "empirical generalization, namely that SIs can occur systematically and freely in arbitrarily embedded

---

8. The reason why 5% and 6% sum to 10% is that all percentages are rounded to the nearest integer.

positions". Our data suggest Chierchia et al.'s "empirical generalization", which by the way is squarely based on introspection, is simply mistaken. Our last experiment markedly failed to confirm that SIs can occur in the scope of quantifiers, and we have argued that, amongst the remaining scope-bearing expressions studied in this paper, the only one that may have licensed (seemingly) local SIs is "think".

On the Gricean view, scalar implicatures can never arise within the scope of any expression. However, this view allows for the possibility that *seemingly* embedded SIs occur under special circumstances.[9] Let us illustrate this point with the help of an example discussed at the outset of this paper:

(31)   Bob believes that Anna ate some of the cookies. (= (2))

We have seen that the scalar implicature Gricean theories deliver for this sentence is the following:

(32)   It is not the case that Bob believes that Anna ate all the cookies.

Now suppose that Bob knows whether or not Anna ate all the cookies; that is, the following is true:

(33)   Either Bob believes that Anna ate all the cookies or he believes that she didn't.

In some contexts, this will be part of the common ground between the interlocutors, or at least a plausible assumption for the audience to make, while in other contexts it may be more dubious. But whenever it is assumed that (33) is true, then it follows from (32) that:

(34)   Bob believes that Anna didn't eat all the cookies.

Though it may seem as if a SI was derived within the scope of "believe", on the Gricean view that is really an illusion: (34) just follows from the implicature shown in (32), provided (31) is uttered in a context in which (33) holds.

An important feature of this explanation is that it doesn't readily generalise to other forms of embedding. To see this, compare the foregoing

---

9. See Recanati (2003), Sauerland (2004), van Rooij and Schulz (2004), Geurts (2006, 2009), Russell (2006), and Spector (2006). The following analysis of belief reports is due to van Rooij and Schulz (2004) and Russell (2006).

example to:

(35)   All the customers shot at some of the salesmen.

For this sentence the Gricean account yields the following implicature:

(36)   Not all the customers shot at all the salesmen.

Conventionalist theories produce a stronger construal:

(37)   All the customers shot at some but not all of the salesmen (hence, none of the customers shot at all the salesmen).

In a Gricean framework we could mimick this result if we could help ourselves to the following auxiliary assumption:

(38)   Either all the customers shot at all the salesmen or none of the customers shot at all the salesmen.

Between them, (36) and (38) entail that none of the customers shot at all the salesmen. However, it will be clear that this assumption is considerably less plausible than (33), and for this reason the Gricean story about belief reports does not generalise to universal quantifiers.

Seen from a Gricean perspective, the general picture about embedded SIs is the opposite from what is to be expected on a conventionalist view. According to Chierchia and his fellow conventionalists, SIs occur "systematically and freely" in embedded positions, which we take to mean that, whenever a SI fails to show up in an embedded position, conventionalists will have to argue that special circumstances apply. Gricean theories, by contrast, rule out the possibility that scalar implicatures occur in embedded positions, which means that, whenever an embedded SI is reported, they are not implicatures (though they may be contingent on implicatures) and special circumstances apply. Such special circumstances may hold in the case of belief reports, as we have seen; other cases may be disjunctions (Sauerland 2004, van Rooij and Schulz 2004), indefinites (van Rooij and Schulz 2004, Geurts 2006, 2009), and presupposition triggers (Geurts 2006, 2009).

The general picture emerging from the foregoing experiments is in line with Gricean expectations. By and large, embedded SIs appear to be rare, if that, and the least unpromising candidacy for the status of being an expression that allows SIs to arise within its scope is by "think" (or "believe");

which is just fine on a Gricean view. Add to this the familiar point that, as a matter of methodological hygiene, pragmatic explanations are to be preferred to lexical or syntactic ones, and it is clear that there is precious little reason for abandoning the Gricean partyline.

Conclusion

Conventionalist theories are designed to explain how scalar inferences can occur in embedded positions. However, the experimental evidence we have mustered indicates that, by and large, scalar inferences do not occur in embedded positions. If this is so, the theory suffers from presupposition failure and in any case it exhibits a remarkable lack of empirical success in its own right, which is only exacerbated by the fact that the standard Gricean approach seems to be doing just spiffingly.

This, then, is one of our two main conclusions. The other is that introspective evidence deserves to be handled with a lot more care than is customary in linguistics and philosophy. We have argued that there are good reasons for being wary of the inference paradigm, which is a straightforward extension of the introspective method of data collection, and have adduced experimental evidence that confirms our suspicion. Introspection has been touted as an exquisite instrument for observing language and meaning, which will only improve with use. However, it may well turn out that the illustrious "trained ear" is actually a potent source of artefactual data.

Acknowledgements

References

Bott, L. and I. A. Noveck (2004). Some utterances are underinformative: the onset and time course of scalar inferences. *Journal of memory and language* 51: 437–457.

Breheny, R., N. Katsos, and J. Williams (2006). Are generalised scalar implicatures generated by default? An on-line investigation into the role of context in generating pragmatic inferences. *Cognition* 100: 434–463.

Chater, N. and M. Oaksford (1999). The probability heuristics model of syllogistic reasoning. *Cognitive psychology* 38: 191–258.

Chierchia, G. (2004). Scalar implicatures, polarity phenomena and the syntax/pragmatics interface. In A. Belletti (Ed.), *Structures and beyond*, pp. 39–103. Oxford University Press.

Chierchia, G. (2006). Broaden your views: implicatures of domain widening and the 'logicality' of language. *Linguistic inquiry* 37: 535–590.

Chierchia, G., D. Fox, and B. Spector (2008). The grammatical view of scalar implicatures and the relationship between semantics and pragmatics. To appear in the *Handbook of semantics*.

Cohen, L. J. (1971). The logical particles of natural language. In Y. Bar-Hillel (Ed.), *Pragmatics of natural language*, pp. 50–68. Dordrecht: Reidel.

Evans, J. S., S. E. Newstead, and R. M. Byrne (1993). *Human reasoning: the psychology of deduction*. Hove, East Sussex: Lawrence Erlbaum.

Fox, D. (2007). Free choice and the theory of scalar implicatures. In U. Sauerland and P. Stateva (Eds.), *Presupposition and implicature in compositional semantics*, pp. 71–120. Houndmills, Basingstoke: Palgrave Macmillan.

Geurts, B. (1999). *Presuppositions and pronouns*. Oxford: Elsevier.

Geurts, B. (2006). Implicature as a discourse phenomenon. In E. Puig-Waldmüller (Ed.), *Proceedings of Sinn und Bedeutung* 11, pp. 261–275. Barcelona: Universitat Pompeu Fabra.

Geurts, B. (2009). Scalar implicature and local pragmatics. *Mind and language* 24: 51–79.

Geurts, B. and F. van der Slik (2005). Monotonicity and processing load. *Journal of semantics* 22: 97–117.

Horn, L. R. (1972). *On the semantic properties of the logical operators in English*. Ph.D. thesis, University of California at Los Angeles.

Horn, L. R. (1989). *A natural history of negation*. Chicago University Press.

Landman, F. (1998). Plurals and maximalization. In S. Rothstein (Ed.), *Events and grammar*, pp. 237–271. Dordrecht: Kluwer.

Landman, F. (2000). *Events and plurality: the Jerusalem lectures*. Dordrecht: Kluwer.

Levinson, S. C. (2000). *Presumptive meanings*. Cambridge, Massachusetts: MIT Press.

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

Noveck, I. and D. Sperber (2007). The why and how of experimental pragmatics: the case of 'scalar inferences'. In N. Burton-Roberts (Ed.), *Pragmatics*, pp. 184–212. Palgrave Macmillan.

Pouscoulous, N. (2006). *Processing scalar inferences*. Doctoral dissertation, EHESS, Paris.

Pouscoulous, N., I. Noveck, G. Politzer, and A. Bastide (2007). A developmental investigation of processing costs in implicature production. *Language acquisition* 14: 347–376.

Recanati, F. (2003). Embedded implicatures. *Philosophical perspectives* 17: 299–332.

Russell, B. (2006). Against grammatical computation of scalar implicatures. *Journal of semantics* 23: 361–382.

Sauerland, U. (2004). Scalar implicatures in complex sentences. *Linguistics and philosophy* 27: 367–391.

Spector, B. (2006). *Aspects de la pragmatique des opérateurs logiques*. Doctoral dissertation, University of Paris VII.

van Rooij, R. and K. Schulz (2004). Exhaustive interpretation of complex sentences. *Journal of logic, language and information* 13: 491–519.

Appendix: Sample materials used in Experiments 1a-b

(1) Betty thinks that Fred has read some of the Harry Potter books.
    $\overset{?}{\leadsto}$ Betty thinks that Fred hasn't read all the Harry Potter books.

(2) Betty thinks that Fred has bought some of Daniel Pennac's novels.
    $\overset{?}{\leadsto}$ Betty thinks that Fred hasn't bought all of Daniel Pennac's novels.

(3) Betty wants Fred to listen to some Beethoven symphonies.
    $\overset{?}{\leadsto}$ Betty wants Fred not to listen to all Beethoven symphonies.

(4) Betty wants Fred to rent some Tarantino movies.
    $\overset{?}{\leadsto}$ Betty wants Fred not to rent all Tarantino movies.

(5) Fred has to read some poems from *Fleurs du Mal*.
    $\overset{?}{\leadsto}$ Fred is not allowed to read all poems from *Fleurs du Mal*.

(6) Fred has to see some of the Harry Potter movies.
    $\overset{?}{\leadsto}$ Fred is not allowed to see all the Harry Potter movies.

(7) All the students have seen some of the plays by Racine.
    $\overset{?}{\leadsto}$ None of the students has seen all the plays by Racine.

(8) All the students have heard some of the Verdi operas.
    $\overset{?}{\leadsto}$ None of the students has heard all the Verdi operas.