Experimental evidence for embedded scalar implicatures — *Emmanuel Chemla & Benjamin Spector*

Geurts and Pouscoulous (2009) provide experimental evidence which they interpret as showing that a sentence such as (1) below cannot be interpreted as 'Every student solved some but not all of the problems' (let us call this putative reading of (1) its 'strong reading'). On this basis, G&P argue against so-called 'localist' theories of scalar implicatures (SIs for short), according to which SIs can be computed in embedded positions. We will argue that G&P's data do not warrant such a conclusion, by showing that the strong reading can in fact be detected by experimental means for sentences such as (1) [Exp.1]. However, this finding is not sufficient to vindicate localist theories, because the 'strong reading' is predicted to be possible not only by localist theories, but also by most current formalized theories of Gricean reasoning ('globalist' theories, cf. Spector 2003 van Rooij & Schulz 2004). We will thus also investigate a case where localist theories and globalist theories are bound to make opposite predictions [Exp.2].

(1) Every student solved some of the problems

Limitations of Geurts and Pouscoulous' (2009) methodology. G&P collected truth-value judgments for sentence-picture pairs such as in Fig. 1, asking subjects to evaluate the relevant sentence as true, false, or ambiguous between a true reading and a false reading. They found that virtually all the subjects considered the sentence to be true in Fig. 1 (below), even though it is false under its strong reading (the top square is linked to *all* the circles). However, the fact that a reading is not detected in a particular experimental task does not provide direct evidence that the reading in question does not exist. In fact, there are several reasons why the strong reading, even if it existed, might have been very hard to detect:

– (i) G&P's pictures are hard to decipher; in particular, the unique falsifier of the strong reading (i.e. the top square) is hard to identify.

– (ii) Disambiguation involves considerations of relevance: one normally understands an ambiguous sentence under its most relevant interpretation without much effort, and even without being aware of its other readings. But for the strong reading to be relevant, the context should somehow make salient a question such as 'Which squares are connected to which circles', or 'Are the squares connected to none, some, or all the circles?'.

– (iii) Several recent theories predict that (1) is ambiguous between a) its literal reading, b) a pragmatic reading which entails that not all the squares are connected with *all* the circles (this is what we will call the 'weak reading') and c) the strong reading (which entails that *no* square is connected with all the circles). Now, note that the strong reading a-symmetrically entails the weak reading, which in turn a-symmetrically entails the literal reading. Various works (in particular Meier & Sauerland 2008) have argued that some kind of a charity principle leads subjects to interpret ambiguous sentences under their weakest readings. If this is correct, the strong reading is expected to be hard to detect experimentally even if it exists.

A variant of G&P's experimental paradigm. We also used a sentence-picture matching task, but with some crucial modifications. Fig. 2 presents one of our conditions, which is the counterpart of Fig. 1, used by G&P: the sentence is true under its literal and weak readings, but not under its strong reading.

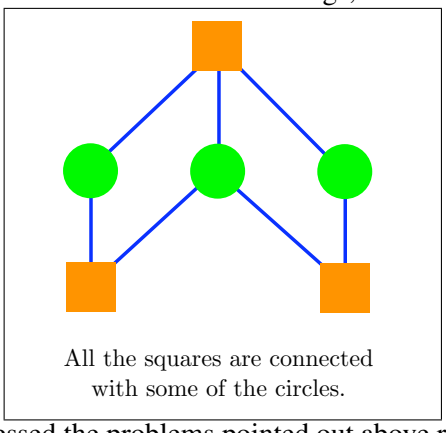


Fig. 1: Item from G&P

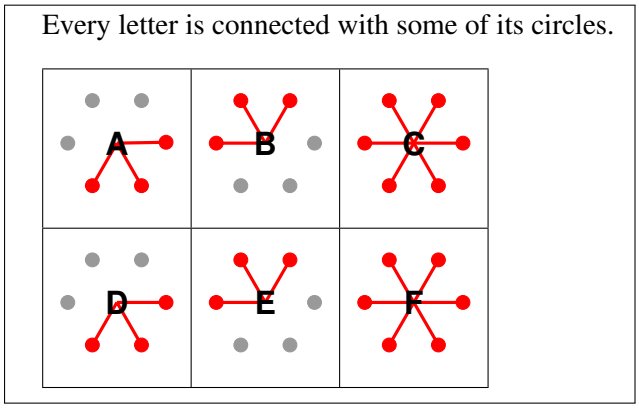


Fig. 2: Corresponding item of ours

We addressed the problems pointed out above regarding G&P's design in the following way:

– (re i) In Fig. 2, the falsifiers of the strong reading are easy to identify (C and F are connected to *all* their circles).

– (re ii) By using *distinct* letters, surrounded by *different* circles, we might be able to draw the subjects' attention to the individual properties of each of them, thus potentially raising the relevance of the strong reading.

– (re iii) Instead of asking for absolute truth-value judgments (see Fig. 3), we asked for *graded* judgments (see Fig. 4): subjects were asked to use a mouse to position a cursor on a line, and were told that the right bound of the line

represented 'true' and the left bound of the line represented 'false'. By collecting more fine-grained judgments than in an absolute truth-value judgment task, we hoped to be able to detect a (possibly unconscious) impact of the strong reading, despite subjects' general tendency to favor weaker readings. Specifically, we speculate that, in the case of an ambiguous sentence, the degree to which it is judged true increases with the number of readings that are true.

Fig. 3: G&P's absolute judgments

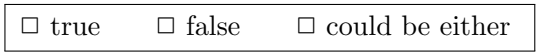


Fig. 4: Our graded judgments

Experiment 1. We tested the sentence 'Every letter is connected with some of its circles' (see fig. 2) (We also tested its counterpart with the scalar item 'or' instead of 'some': 'Every letter is connected with its blue circle or with its red circle', and obtained similar results). We paired this sentence with different pictures, giving rise to the following four target conditions: **FALSE:** no reading is true, **LITERAL:** only the literal reading is true, **WEAK:** both the literal and the weak reading are true but the strong reading is false (as in Fig. 2), and **STRONG:** all readings are true.

The answers given by our 16 participants (see Fig. 5 below) show that the degree to which the sentence is judged true is, on average, significantly higher in the **STRONG** condition than in the **WEAK** condition ($F(1, 15)$=25,$p < .001$). This difference is straightforwardly explained only if the strong reading exists.

Control conditions using downward entailing environments ('no' instead of 'every') showed that subjects did not systematically interpret scalar items like 'some' as meaning 'some but not all'.

Moving to non-monotonic contexts. Consider the following sentence:

(2) There is exactly one letter connected with some of its circles.

In this case, the three potential readings we are interested in are: 1. the literal reading, 2. the 'local' reading: 'There is exactly one letter connected with some but not all of its circles', and 3. the 'global' reading: 'There is exactly one letter connected with some of its circles, and no letter is connected with all of its circles'. The global reading is predicted by most formalized version of Gricean reasoning (cf. Spector 2007). The 'local' reading, however, is predicted only by localist theories, because the local reading fails to entail the literal reading in this case, and globalist theories can only predict readings that entail the literal reading. Note also that the entailment relation between the local and global readings is reversed: the global reading entails the local reading (as well as the literal reading). It follows that a) non-monotonic contexts provide us with a way to distinguish between localist theories and globalist theories, and b) if the local reading exists, it should be easy to detect (since it is not stronger than the other readings). G&P tested such sentences, but concluded that their results were hard to interpret, due to the difficulty of the task.

Experiment 2. In Experiment 2, we tested the sentence in (2) with the same kind of pictures as in Experiment 1, and asked again for graded judgments (counterparts with the scalar item 'or' instead of 'some' were also tested, with similar results). We were interested in four types of conditions: **FALSE:** all potential readings are false, **LOCAL:** the local reading is true but both the literal and the global readings are false, **LITERAL:** the literal reading is true but both the local reading and the global reading are false, and **GLOBAL (ALL):** the global reading is true (and therefore the local and literal readings are true as well).

The answers of our 16 participants (Fig. 6 below) provide clear evidence for the existence of the local reading. In particular, even though the literal reading is false in the **LOCAL** condition, the degree to which the sentence is judged true in this condition is higher that in the **LITERAL** condition ($F(1, 15)$=6.72, $p < .05$). As expected, the highest scores are found with the **GLOBAL** condition (where all readings are true).

Fig. 5&6: Mean position of the cursor in the various target experimental conditions

Fig. 5: Experiment 1

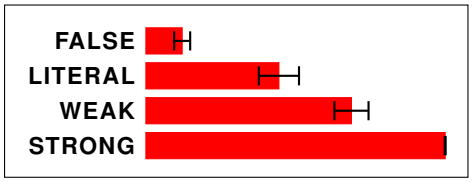
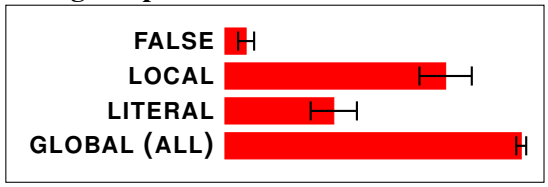


Fig. 6: Experiment 2



Conclusion. Our results suggest that, contrary to G&P's conclusions, sentences such as (1) can be interpreted under what we called the 'strong' reading, which is consistent both with localist theories and a subclass of globalist theories. Futhermore, it appears that scalar items can be interpreted under their strong meaning when they occur in a non-monotonic context, even though the resulting reading is logically independent of the literal reading. This finding provides direct experimental evidence for embedded SIs.