# The *best response paradigm*: A new paradigm to test implicatures of complex sentences

## Introduction

It is a controversial theoretical debate whether implicatures can arise in embedded positions. To give an example, researchers disagree whether the sentence *Each girl found some of her marbles* carries the strong inference that each girl found some but not all of her marbles (e.g., Sauerland, 2004; van Rooij & Schulz, 2004; Chierchia, 2004; 2006; 2013). For the given example, globalists predict a weaker implicature that not all girls found all marbles (e.g., Sauerland, 2004). Hence, the question concerning the availability of embedded implicatures provides an important test bed adjudicating between different theories of implicature.

Experimental research concerning the existence of embedded implicatures has yielded contradictory results (Geurts & Pouscoulous, 2009; Chemla & Spector, 2011). The current literature disagrees on what exactly different experimental paradigms measure and whether previous findings might be heavily based on typicality effects rather than the computation of embedded implicatures *per se* (e.g., Geurts & van Tiel, 2013; van Tiel, 2014).

Here, we present clear experimental evidence for embedded implicatures in a novel experimental paradigm. Our *best response paradigm* provides an organic setting which (i) makes the derivation of implicatures relevant and (ii) is less prone to typicality effects. We further compare our novel paradigm to the sentence-picture verification paradigm used in previous studies.

## Best response paradigm

*Methods*: The scenario we used for our best response paradigm is an extension of the marbles paradigm by Degen & Goodman (2014). We included a reward system to make the contrast between *some* and *all* relevant for the task participants have to perform. In this scenario, there are four sisters each owning a set of four special edition marbles, which get lost while the girls are playing. Participants are told that the mother of the girls wants to reward her girls depending on how many marbles they find. A girl gets (i) chocolate if she finds all 4 of her marbles, (ii) candy if she finds fewer than 4 of her marbles and (iii) a gummy bear when she finds none of her 4 marbles (as a consolidation prize). The task of the participants is to buy sweets for the four girls depending on the statements the mother utters. For example, if the mother says *No girl found any of her marbles* participants should only buy gummy bears. Participants were asked to give binary responses (yes/no) for each of the three types of sweets: chocolate, candy and gummy bears. We used two test sentences in which *some* was embedded (1) under *each*, in which case localism makes a stronger prediction than globalism and (2) under *some*, where globalism makes a stronger prediction than the version of localism presented in Chierchia (2004). Table 1 summarizes the critical sentences and the predicted response patterns according to inference participants compute.

Table 1: Overview of conditions and predictions

| Quantifier condition | Inference | Chocolate | Candy | Gummy bear |
|---|---|---|---|---|
| (1) Each girl found | localist (strong: $\neg\exists\forall$) | no | yes | no |
| some of her marbles | globalist (weak: $\neg\forall\forall$) | yes | yes | no |
| (2) Some of the girls found | localist* (weak: $\neg\forall\exists$) | yes | yes | yes |
| some of their marbles | globalist (strong: $\neg\exists\forall$) | no | yes | yes |

The two critical conditions were repeated twice. We further included a control item with the sentence *Each girl found some and possibly all marbles* as well as seven additional filler items (for example the sentence *Each girl found all marbles*). Hence, each participants saw 12 items in total.

*Results*: 40 native English speakers were recruited on Amazon's Mturk platform. All participants gave consistent responses in the filler items. Figure 1 shows the mean percentage of YES responses across critical conditions for each reward type. If participants do not buy chocolate, they

have computed the strong inference that no girl found all of her marbles. In the condition with each_some 4 % of the participants bought chocolate and 12% in the case of some_some.

We computed a logit mixed effects model including the fixed factors quantifier condition (each_some vs. some_some), reward type (bear, candy, chocolate) and their interaction (reference level: each_some, candy). The model showed a significant difference across reward type: participants chose candy ('some') more often than chocolate ('all', p <.001) and than bears ('none', p <.001), corresponding to the reading that each girl found some but not all of her marbles. There was no overall difference between the two quantifier conditions. Importantly, the model revealed an interaction between quantifier condition and reward type in that participants chose bears more often in the condition with some_some (bear vs. candy across quantifier condition: p <.01). This response pattern is consistent with the reading that no girl found all marbles and some did not find any marbles in the some_some condition. The control item showed that participants bought both candy and chocolate (92 %) when they were uncertain whether a girl found some or possibly all marbles.
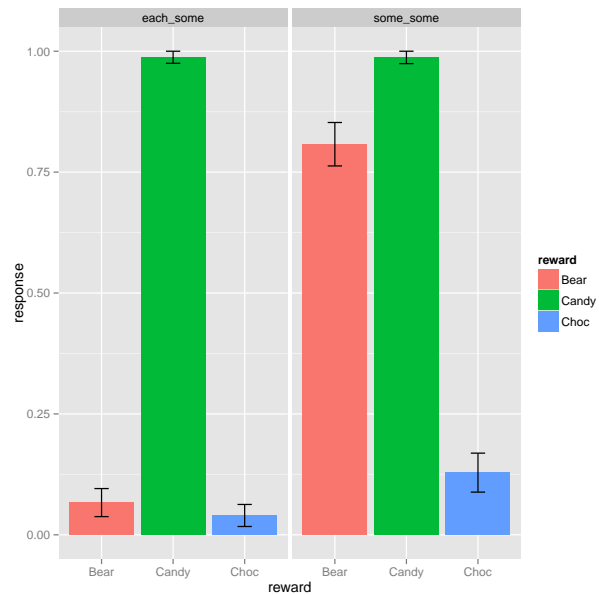


Figure 1: % YES for every_some and some_some across reward (bears 'none' = red, candy 'some' = green, chocolate 'all' = blue)

*Conclusions*: The results of the study provide clear evidence that participants compute the strong inference that no girl found all marbles and that this reading is the preferred interpretation. The strong inference for the sentence *Some of the girls found some of the marbles* is not predicted by Chierchia (2004) but in more recent versions of the localist account. In summary, our paradigm yields a high rate of embedded implicatures in a scenario where the different readings of the sentence are relevant for the response.

### Geurts & Pouscoulous (2009, G&P) and Chemla & Spector (2011, C&S)

In the remainder, we will discuss the reasons for the contradictory findings in previous experiments. We ran a series of follow up studies on the paradigms by G&P and C&S. In the paradigm by G&P we did not observe the weak globalist inference and a low number of unembedded implicatures. Hence, the paradigm yields a floor effect. In C&S's paradigm participants differentiate the weak and strong inference but only after they have seen the strong condition (see van Tiel, 2014).

### Conclusions

Our novel best response paradigm provides clear evidence for embedded implicatures in a setting that mirrors natural language comprehension and we did not ask participants for metalinguistic judgements. Contrary to previous studies, the results presented here cannot be explained by typicality. We further found evidence for a reading that is not predicted by the theory of localism presented in Chierchia (2004). In future work, we apply our paradigm to other relevant cases such as non-monotone contexts in order to adjudicate between different theories of implicature.

**Selected References:** • **Chemla, E. and Spector, B. 2011**. Experimental evidence for embedded scalar implicatures *Journal of Semantics*.