Why some children accept under-informative utterances: Lack of competence or Pragmatic Tolerance?

Alma Veenstra, University of Cambridge, UK

Bart Hollebrandse, University of Groningen, Netherlands

Napoleon Katsos, University of Cambridge, UK

Abstract

Binary judgement on under-informative utterances (e.g. *Some horses jumped over the fence*, when all horses did) is the most widely used methodology to test children's ability to generate implicatures. Accepting under-informative utterances is considered a failure to generate implicatures. We present off-line and reaction time evidence for the *Pragmatic Tolerance Hypothesis*, according to which some children who accept under-informative utterances are in fact competent with implicature but do not consider pragmatic violations grave enough to reject the critical utterance. Seventy-five Dutch-speaking four to nine-year-olds completed a binary (Experiment A) and a ternary judgement task (Experiment B). Half of the children who accepted an utterance in Experiment A penalised it in Experiment B. Reaction times revealed that these children experienced a slow-down in the critical utterances in Experiment A, suggesting that they detected the pragmatic violation even though they did not reject it. We propose that binary judgement tasks systematically underestimate children's competence with pragmatics.

Keywords: implicature, pragmatic development, under-informative utterances, reaction times

1. Introduction

According to Grice (1975 [1989]), interlocutors are expected to adhere to the Cooperative Principle and a set of maxims that makes their conversation as efficient as possible. Cooperative speakers will comply with the Maxims of Quantity, Quality, Relation, and Manner, which enjoin them to say no more and no less than what is needed, to be truthful, relevant, concise and unambiguous. In this paper we will focus on the development of children's competence with the Maxim of Quantity and especially its first sub-maxim, which prescribes speakers to make their contribution as informative as is required.

Numerous studies in the field of pragmatics report that children are less competent using this maxim to derive pragmatic inferences (known as *implicatures*) than adults. In a high-profile paper that summarizes the emerging findings, Noveck and Reboul (2008) describe the development of informativeness as moving from broad semantic interpretations to narrowed pragmatic interpretations. The statement *some apples are in the basket* simply means that two or more apples are in the basket. A narrowed interpretation based on the Maxim of Quantity involves the evaluation of the speaker's intentions and of other, more informative options: what was said, and what else could have been said? If all apples were in the basket, then a cooperative speaker would have stated the more informative proposition by saying *all apples are in the basket*. Therefore, the fact that they did not use the more informative proposition (assuming that they could have done so) suggests that this is not the case. In the case at hand, the hearer may reason that by only saying *some apples* the speaker also implies that 'not all' of the apples are in the basket. This interpretation is typically known as a *quantity implicature*. Moreover, because the implicature involves expressions that can form a scale of informativeness (like '*some* and *all*), it is further known as a *scalar implicature* (see Geurts 2010).

In terms of child development, Noveck (2001) used a binary judgement paradigm, where participants are presented with an under-informative statement and are asked to make a binary judgement in terms of truth or falsity (or correctness, in other studies). Noveck (2001) found that 8- and 10-year-old children were not rejecting a sentence with a weaker term on a scale in favour of the more informative term. For example, the children judged sentences like *some giraffes have long necks* as true, despite the fact that all giraffes have long necks. Importantly, these children did not have problems rejecting sentences in general, as they were adult-like at rejecting sentences that were logically false. Papafragou and Musolino (2003), who also used a binary judgement paradigm, tested 5-year-old children's performance on quantified, numerical, and verbal scales, and found that children predominantly accepted under-informative statements. Training and explicit instruction improved the children's performance, increasing the success on the numerical scales (but not the quantified or verbal ones) to a near-ceiling 90%. Feeney and colleagues (2004) found that manipulation of the relevance of the implicature also enhances rejections in 7-year-olds. Guasti and colleagues (2005) studied children's pragmatic competence further in line with Papafragou and Musolino's (2003) training and explicit instructions. They found that the performance of 7-year-olds did improve, but that the effect did not persist over a longer period of time. Barner, Brooks and Bale (2011) and Papafragou and Tantalou (2004) also used binary judgements and further demonstrated that children's performance does not differ when other types of under-informative sentences are tested (e.g., when the sentence *the cow and the dog are sleeping* describes a situation where a cat is sleeping as well). Foppolo, Guasti and Chierchia (2012) demonstrated the role of different lexical choices (partitive vs. non-partitive quantifiers) and the salience of the more informative expression that was not used (e.g. *all*), while Skordos and Papafragou (2016) demonstrated the role of contextual relevance of the more informative expression, in both cases by studying children's rates of rejection of under-informative utterances.

These and other papers have significantly advanced our understanding of the factors that are relevant when children (and adults) decide how to judge the felicity of an under-informative utterance. What is important for the present investigation is that these studies base their conclusions on the assumption that the children who accept under-informative utterances altogether lack (some or all of the) competence that is needed to generate implicatures. It is beyond the scope of this paper to review exactly what kind of competence children who accept under-informativeness lack (this is a theory-critical matter which has generated a lot of attention and some controversy, see Katsos 2014). What we highlight is that they all assume that these participants lack some kind of competence and that this prevents them from generating implicatures.

However, Davies and Katsos (2010), Katsos and Smith (2010), and Katsos and Bishop (2008, 2011) have argued that children's acceptance of under-informative sentences may have a different origin. Katsos and Bishop (2011) used a ternary task in addition to a binary task, and found that children were much more reluctant to endorse under-informative sentences in the former compared to the latter. That is, while 5-year-old children accepted under-informative utterances when given a binary choice (to decide if the utterance was right or wrong), they penalised the utterances in a ternary scale (to decide whether to award a small, medium or large strawberry), by giving the medium or the small rather than the top award. This finding gave rise to the *Pragmatic Tolerance Hypothesis*: children are sensitive rather than blind to violations of the Maxim of Quantity, but they do not always take the violation as severe enough to warrant a downright rejection of the critical sentence. Once children are tested in a paradigm that does not require them to demonstrate their competence by the categorical rejection of pragmatically infelicitous sentences, then children's true competence may be evident at ages much younger than the ones tested up to now.

Katsos and Bishop (2011), and Katsos and Smith (2010), who reached similar conclusions, tested two different groups of 5-year-old children: one that was given a binary judgement task, and one that was given a ternary judgement task. Davies and Katsos (2010) tested sensitivity to over- and under-informativeness with a binary and a graded judgement task **within** the same groups of participants, and showed that children who seemed insensitive in the binary task did display sensitivity in the graded task.

In this paper, we have two aims: First, we aim to replicate the evidence in support of the Pragmatic Tolerance Hypothesis by using a binary and graded judgement task within the same group of participants with a new set of stimuli, ad hoc scales. We propose the developmental hypothesis that children are competent with implicature, and by extension, with some crucial aspects of Gricean pragmatics, at a much younger age than previously considered. This is anticipated by several studies in word-learning and mutual exclusivity, which propose that children as young as two years old employ Gricean-like principles to infer the likely referent of a novel word (see Clark 1990, among others; for an opposite view of children's competence with mutual exclusivity, see De Marchena et al. 2011).

Second, we wish to advance our understanding of the process through which pragmatically competent children tolerate pragmatically under-informative sentences. Currently, there is ample evidence that when participants are rejecting under-informative sentences, they require more time to do so than when rejecting logically false ones. Relevant data were first presented by Noveck and Posada (2003), who looked at reaction times to under-informative statements in adults. Similar data are reported by Bott and Noveck (2004) and De Neys and Schaeken (2007). The studies by Noveck and colleagues also report that rejecting under-informative sentences tends to take more time than accepting them.

In contrast, Feeney and colleagues (2004) found longer reaction times for adult participants accepting under-informative sentences than rejecting them . They claim that this longer reaction time is due to the reflective process of generating the pragmatic interpretation of an utterance and then inhibiting it to revert to the logical interpretation. A positive correlation between acceptances of under-informativeness and success in a counting span task led them to conclude that sufficient cognitive capacity is needed to cancel the implicature.

Thus, the findings from the adult studies seem to be in direct contrast with each other, or to potentially stem from different sources. As discussed in Noveck and Reboul (2008), the studies by Noveck and colleagues as well as the dual task experiment in the De Neys and Schaeken (2007) suggest that in terms of working memory, generating pragmatic inferences is a resource-consuming process. In contrast, the studies by Feeney and colleagues suggest that it is the reflective un-doing of pragmatic inferences that requires superior working memory resources.

Where does Pragmatic Tolerance fit in with these hypotheses? First of all, the major prediction of the Pragmatic Tolerance Hypothesis is that, even when pragmatically competent participants accept under-informative utterances, their acceptance should take more time than the acceptance of fully informative utterances (e.g. *some elephants have trunks* vs. *some elephants live in zoos*). This delay would be a reflection of the fact that participants have detected that the under-informative sentences violates the Maxim of Quantity, whereas no such violation is present in the fully informative cases. This prediction has not been the focus of previous work, and it is the minimum one would expect if the hypothesis on Pragmatic Tolerance is on the right track.

This leads to different predictions for children who are pragmatically tolerant than for children who are indeed not competent with implicature. We have predicted that children who are not competent with implicature altogether will accept under-informative sentences in a binary judgement task. They will also rate these sentences optimally in a ternary paradigm, since to them under-informative sentences are as felicitous as straightforwardly pragmatically appropriate sentences. Consequently, the reaction times for judging under-informative and fully informative sentences in the binary judgement task will be comparable. For pragmatically tolerant children, on the other hand, who accept under-informativeness in a binary setting, we predict that they show sensitivity to under-informativeness in a ternary judgement task by not awarding the top reward. Moreover, the reaction times in the binary judgement task ought to be longer for accepting under-informative sentences as compared to accepting logically true sentences (which would reflect the fact that they detected that the sentence was infelicitous, even though they accepted the sentence in the binary judgement task). These predictions are shown in Table 1.

Table 1. Predictions of behaviour towards under-informative statements.

	Childrens' competence with implicature	**Binary Task**	**Ternary Task**	**RT in the Binary Task**
1	Not competent in some respect	Accept	Top reward	Acceptance of under-informative ≈ Acceptance of fully informative
2	Pragmatically tolerant	Accept	Non-top reward	Acceptance of under-informative > Acceptance of fully informative
3	Pragmatically competent	Reject	Non-top reward	Rejection of under-informative > Rejection of false utterance
4a	Pragmatically competent	Reject	Non-top reward	Rejection of under-informative > Acceptance of under-informative
4b	Pragmatically competent	Reject	Non-top reward	Acceptance of under-informative > Rejection of under-informative

Note: Predictions 1–3 stem from the Pragmatic Tolerance Hypothesis. Prediction 4a follow Bott and Noveck (2004), Noveck and Possada (2003), and De Neys and Schaeken (2004), whereas 4b is predicted by Feeney et al. (2004).

In the current study, informativeness was operationalised in a design where different objects or animals were inside a basket or on a sofa. This allowed us to create fully informative or under-informative sentences by using the same linguistic materials (noun phrases), but varying whether the sentence made mention of just one or all of the objects or animals. Therefore, under-informative utterances (e.g. *in the basket there is a shoe* in a display with a shoe and a ball inside a basket) would give rise to quantity implicatures based on an ad hoc scale, of the sort tested in Katsos and Smith (2010), Katsos and Bishop (2011), and Barner, Brooks and Bale (2011).

2. Method

2.1 Participants

Seventy-five neuro-typical children from primary school “de Sprankel” in Stiens, the Netherlands, took part in this study. The children were all native speakers of Dutch, and they received stickers for their participation. Their ages range between 4;0 and 9;8 (μ = 6;3). There were 32 boys.

2.2 Materials and Procedure

2.2.1 Experiment A

The binary judgement task was designed in *E-Prime* and administered on a Sony laptop. First, the participants are introduced to a game by a fictional character who explains that the participants are going to see pictures which she will try to describe in the best possible way. The pictures show a basket with several objects inside and outside of it. The object were: *bal* (ball), *schoen* (shoe), *bloem* (flower), *beer* (bear), *schaar* (scissors), *trein* (train), *pet* (cap), *kat* (cat), *banaan* (banana), *boek* (book), *hoed* (hat), and *klok* (clock).

Except for the teddy bear, banana, and cat, used only in the practice items, the object labels were all monosyllabic and the entities depicted were all inanimate. The participants were asked to indicate on a SR serial response box whether or not they thought the fictional cartoon character described the picture correctly. They had to press the green button (correct) or the red button (incorrect) as quickly as possible. Responses and reaction times were registered, with reaction times being measured from the point where the pre-recorded sentence ended until one of the buttons was pushed. While the sentence was played, an empty basket was shown on the screen. The corresponding picture was presented immediately after the sentence, so that the reaction times measured the exact time it took to look at the objects and make a decision about the sentence. Pictures and recorded sentences were divided into five conditions (see Table 2). There were eight items in each condition, which were presented in random order.

Table 2. Conditions in the binary judgement task

	In the basket, there is a…	
	*shoe*	True (1 object)
	ball	False (1 object)
	*shoe and a ball*	True (2 objects)
	shoe and a hat	False (2 objects)
	ball	Under-informative

2.2.2 Experiment B

The ternary judgement task was created on the same principles as the binary judgement task. The participants were introduced to the game by the same girl. She informed the participants that she liked strawberries and that in this game, the participants should reward her with a small strawberry, a medium-sized strawberry, or a large one, in correspondence to how well they thought her sentence described the picture. The participants should always give her at least the smallest strawberry. This 3-point rating scale (see also Katsos & Bishop 2011) is presented in Figure 1.

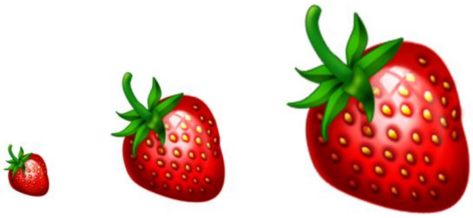

Figure 1. The 3-point rating scale from the ternary judgement task

The pictures in the graded judgement task showed a sofa on which various cartoon animals were sitting. Except for the rabbit and the two deer used only in the practice items, all animal labels were monosyllabic: *konijn* (rabbit), *hertjes* (deer, plural), *geit* (goat), *gans* (goose), *slang* (snake), *muis* (mouse), *kat* (cat), *beer* (bear), *leeuw* (lion), *hond* (dog), *aap* (monkey), *kip* (chicken), and *koe* (cow). All materials are available from the authors upon request.

The participants had to indicate which strawberry they would reward on an SR serial response box, where the leftmost button corresponded with the small strawberry, the middle button with the medium one, and rightmost button with the large strawberry. An example sheet showing the strawberries (Figure 1) was always visible. The items were again divided into five conditions, similar to the binary judgement task. There were eight items in each condition, which were presented at random.

Testing took place in a quiet room at the children's school. All participants started with the binary judgement task, which took about five minutes. The children then performed another set of tasks that was part of another investigation and lasted around 20 minutes. This interval was intended to take the children's attention away from the specifics of the task. After the interval, the participants continued with the graded judgement task, which took also about five minutes. The session took in total around 30 minutes, after which the children were rewarded with a sticker for each different task.

3. Results

3.1 Behavioural results

3.1.1 Experiment A

In the binary judgement task, the participants had to press either a green or a red button as quickly as possible when they thought the sentence was right or wrong, respectively. The children were more or less at ceiling for accepting the true and rejecting the false conditions, see Figure 2:

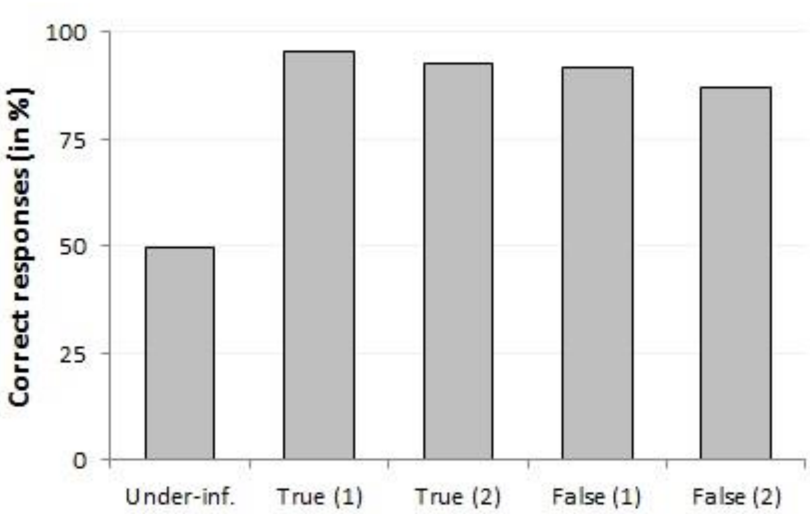

Figure 2. Percentage of correct responses in the different conditions. The number in parentheses refers to the number of objects in the display. Note that in this paper, we label the rejection of an under-informative statement as a correct response.

Of all under-informative items, 50.3% were accepted, with a large variation (SD = 41.4%). Looking at the number of items that each child accepted (out of 8), we see that the majority either never accepted under-informative sentences, or always accepted them, see Figure 3:

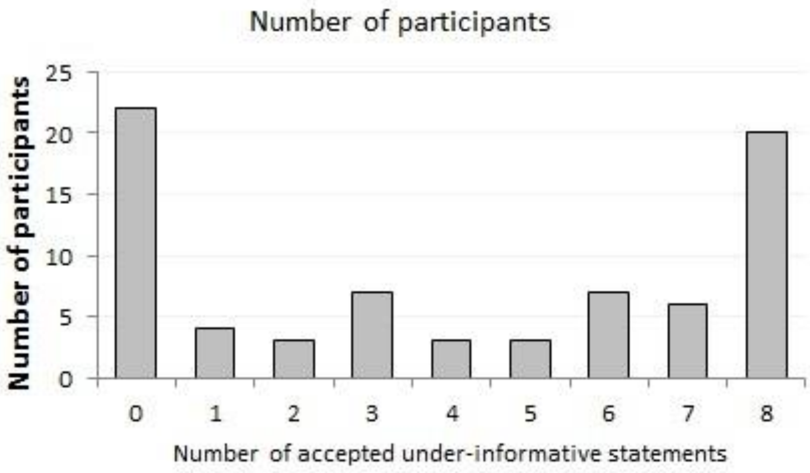

Figure 3. Distribution of acceptances of under-informative statements

Out of the 75 participants, 22 participants never accepted under-informative statements, whereas 20 participants always accepted them. Based on the results from the binary judgement task, one could conclude that there is a clear-cut distinction between children, with some being consistently competent with implicature (n = 22), some being in transition (n = 33), and some consistently (apparently) lacking competence with implicature (n = 20). Such interpretations for this kind of data are consistently found in the literature (see, e.g., Guasti et al. 2005).

3.1.2 Experiment B

In the graded judgement task, the same participants rewarded sentences with a large, medium, or small strawberry (see Figure 4).

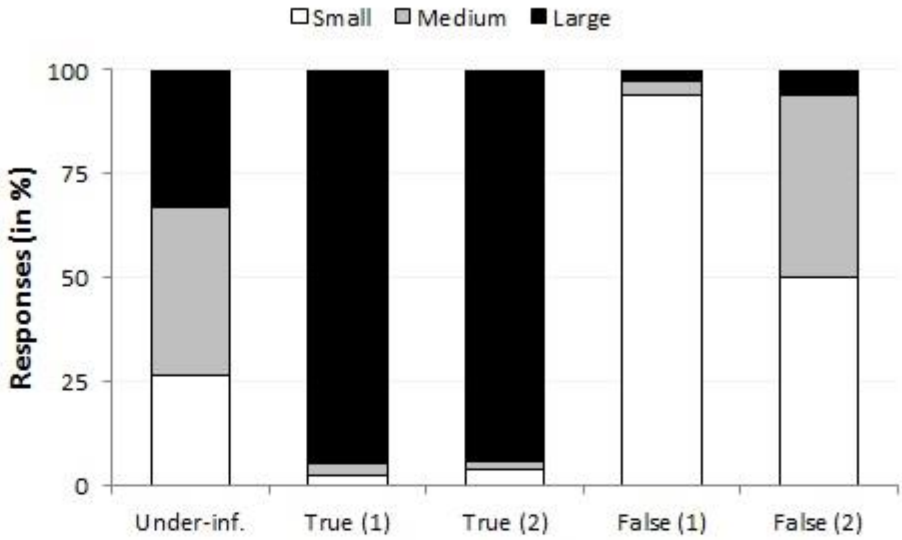

Figure 4. The distribution of small, medium, and large strawberries awarded in each of the different conditions. The number in parentheses refers to the number of objects in the display.

True statements in the (1) or (2) items conditions were rewarded mostly with the top reward (large strawberry), whereas false statements in the (1) or (2) items conditions almost never received the top reward. However, no clear strategy is seen for the under-informative condition. An important question is how the twenty (allegedly not competent) participants who accepted the under-informative statements in the binary judgement task in Experiment A performed in the ternary judgement task. Figure 5 shows that of these 20 children, seven children gave the top reward as a reward for all under-informative sentences, whereas six children never awarded a top reward.

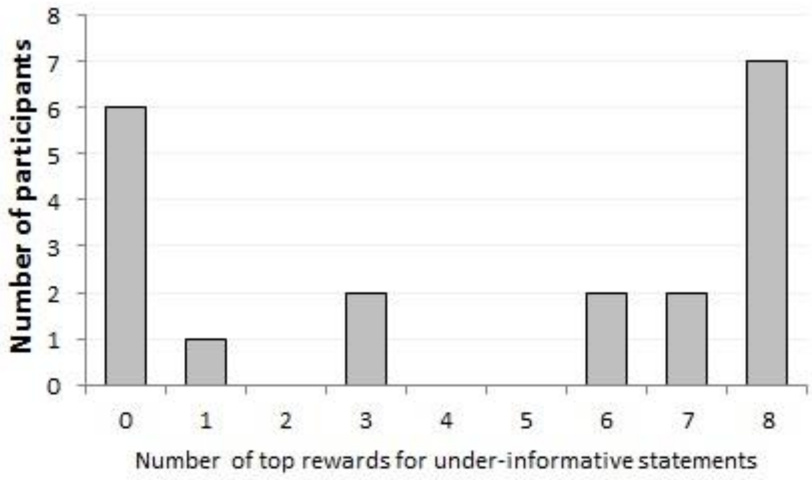

Figure 5. Distribution of under-informative statements that were rewarded with the top reward (large strawberry) by the twenty participants who accepted all statements in the binary task

These results indicate that seven participants in our study were oblivious to pragmatics (and four more participants also mostly gave the top reward for under-informative statements). However, six participants never gave the top reward (and one more almost never gave the top reward), despite accepting under-informativeness in the binary judgement task. These latter children would have been (mis-)classified as lacking competence with implicature by the binary task alone, but show in the ternary task that they are actually able to detect and penalise pragmatic infelicity.

3.2 Reaction times

In the binary judgement task, Experiment A, reaction times (RTs) were recorded for the judgement. Responses that took longer than 10000 ms were removed (2.4% of the data). Overall RTs are shown in Figure 6.

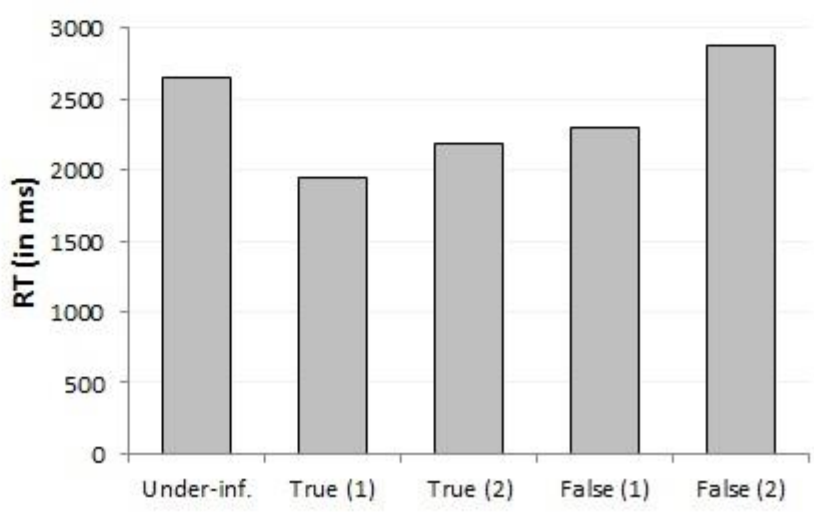

Figure 6. Average response times in the different conditions. The number in parentheses refers to the number of objects in the display.

In general, participants took longer to respond to under-informative statements compared to true statements with one object (692 ms, $t(1172) = 7.424$, $p < .001$). They also took longer to respond to under-informative statements compared to false statements with one object (348 ms, $t(1132{,}135) = 3.735$, $p < .001$). In addition, participants were faster to respond to true statements compared to false statements, both for single object displays (344 ms, $t(1191) = -4.025$, $p < .001$) and double object displays (694 ms, $t(1156) = -6.962$, $p < .001$). Also, displays with one object were responded to faster compared to displays with two objects, both for true statements (238 ms, $t(1181) = -2.8$, $p < .01$) and false statements (588 ms, $t(1166) = -5.889$, $p < .001$).

The judgement data from Experiments A and B combined showed that some children were pragmatically tolerant rather than lacking competence. So first, we compare the RTs for under-informative items to true items (with one object) between the two groups (see Table 3). Note that the pragmatically tolerant group only includes children who accepted all under-informative statements in the binary task, but never awarded them the top reward, a large strawberry. The non-competent group only includes children who accepted all under-informative statements in the binary task and always awarded them the top reward, a large strawberry in the ternary task.

Table 3. Reaction times for the accepters of under-informative statements

	Oblivious children	**Pragmatically tolerant children**
Mean RT for under-informative	2207 ms	2098 ms
Mean RT for true (1 object)	2136 ms	1389 ms
Independent samples *t*-test	$t(99) = 0.21$; $p = .83$	$t(54,6) = 3.05$; $p < .01$

Note: For oblivious children, n = 7, for pragmatically tolerant children, n = 6.

Table 3 shows that there was no difference in RT for the non-competent children between under-informative sentences and true sentences (with one object), confirming Prediction 1 (see Table 1). In contrast, the pragmatically tolerant children showed a delay of on average 709 ms in responding to under-informative sentences compared to true ones, confirming Prediction 2. This pattern is similar to that of the 22 children who were pragmatically competent in the binary task: they showed a 698 ms delay for under-informative sentences compared to true ones ($t(283.9) = 4.347$; $p < .001$), but note that they rejected the under-informative sentences, rather than accepting them.

Next, we compare the RTs for rejecting under-informative sentences with the RTs for rejecting false sentences (with one object) across pragmatically competent children. Rejecting an under-informative sentence took on average 2529 ms, whereas rejecting a false sentence (with one object) took on average 2191 ms. This delay was significant (338 ms, $t(315.131) = 2.181$, $p < .05$), confirming Prediction 3.

Finally, we compare the RTs for accepting under-informative sentences by pragmatically tolerant children with the RTs for rejecting under-informative sentences by the same children. Accepting under-informative statements took on average 2098 ms, whereas rejecting them took on average 2519 ms, but this delay in rejection was not significant (420 ms; $t(192) = 1.65$; $p = .10$), thus not confirming either Prediction 4a or 4b.

4. Discussion

The first objective of this study was to test whether evidence for Pragmatic Tolerance can be found with ad hoc scales within the same subjects. Is there a risk that a child will be labeled as lacking competence with quantity implicature in a binary judgement task, whereas that same child does show sensitivity in a ternary judgement task? Earlier research by Katsos and colleagues (2008; 2010; 2011) suggested that this might be the case; yet, the hypothesis was tested on ad hoc scales **between** subjects, rather than **within** subjects. Our study shows that when the same children are tested in both paradigms, it turns out that some are pragmatically tolerant to under-informativeness and accept under-informative sentences in the binary task, but do not award them the highest reward in the ternary task. Out of the twenty children who accepted under-informative sentences, six were pragmatically tolerant, seven were indeed lacking competence, and seven more were in transition.

The second objective of this study was to test the predictions that the Pragmatic Tolerance Hypothesis makes with regard to reaction times. This hypothesis proposes that while pragmatically tolerant children are competent with pragmatics, they do not think that violations of the Quantity Maxim are severe enough to reject the sentence. We therefore predicted that pragmatically tolerant children take longer to respond to under-informative sentences (for which an inference has to be made) compared to logically true sentences (for which no additional inference is needed). Indeed, we found a significant delay for accepting under-informative sentences. We also predicted that children who lack competence do not differentiate between under-informative sentences (which are logically correct) and informative, true sentences. In line with the predictions, these RTs did not differ.

We also investigated whether it is the generation of an implicature or rather the cancellation of an implicature that requires the most time and resources. If the generation of an implicature is the most costly, rejections of under-informative sentences should be delayed compared to acceptances. On the other hand, if the cancellation of an implicature is the most costly, acceptances of under-informative sentences should be delayed compared to rejections. As our results showed no difference between RTs of rejections and acceptances, the current study cannot answer this question.

Although most of the control conditions in our study yielded clear results, the false condition with two objects seemed to behave differently: it had the highest error rates in the binary task, received many more medium-sized strawberries than the other false condition in the ternary task, and yielded the longest RTs. This pattern was mainly due to the relationship between the sentence and the picture: the picture shows a basket with a ball and a shoe, whereas the sentence mentions a shoe and a hat. Although this statement as a conjunction is wrong, part of it is correct (i.e. there is a shoe in the basket). This might lead to confusion (higher error rates and longer RTs) because it is partly true, which explains why it received the medium-sized strawberry quite often. No such confusion was present for the true conditions, which were always completely true, or for the false condition with one object, which was completely false.

The Pragmatic Tolerance Hypothesis assumes a developmental trajectory of pragmatic competence starting from pragmatically oblivious, via pragmatically tolerant, to pragmatically competent. Although the number of children in each group in our study is very small because of our strict categorization (and the differences are not statistically supported), the oblivious children are on average 17 months younger than the tolerant children, see Figure 7:

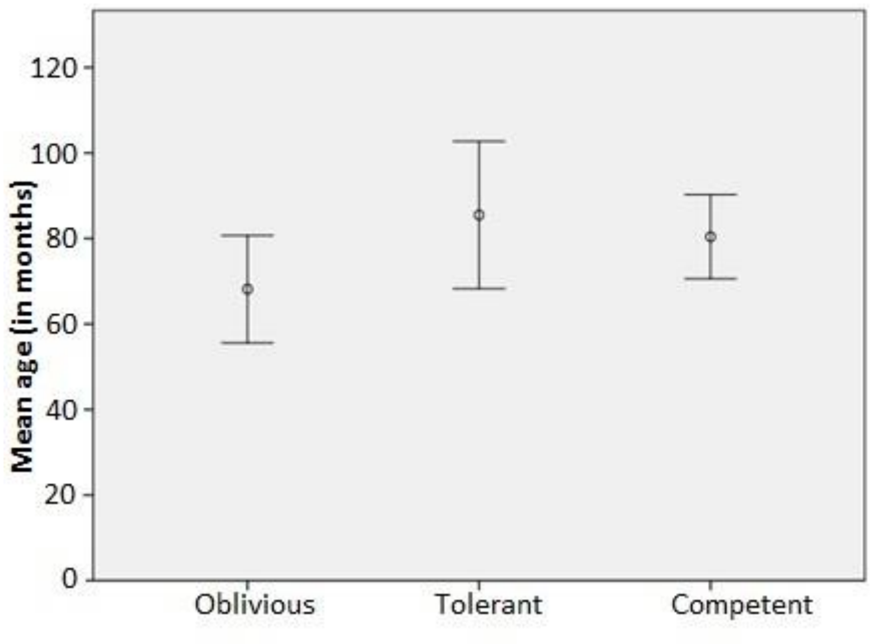

Figure 7. Average age for each of the pragmatic competence categories

Thus far, it is unclear whether each child goes through a pragmatically tolerant stage and why (although there are reports of adults being pragmatically tolerant, we believe those are task effects, rather than a developmental stage in their pragmatic competence, e.g. Katsos & Bishop 2011). If one assumes that pragmatic competence comes with increased cognitive resources, it could be argued that the pragmatically tolerant children use these resources to both derive implicatures and adapt to their still pragmatically oblivious peers who cannot be expected to take into account the Gricean maxims. Another possibility is that these increased resources are still limited, therefore only used when the task demands it (as in a ternary judgement task) and not when a logical answer is sufficient (as in a binary judgement task). Future research might include a longitudinal study to shed more light on the development of pragmatic tolerance.

Finally, we conclude this paper with a word of caution on the use of binary judgement tasks for the study of pragmatic competence. Without singling out any specific study, it is fair to say that to date binary judgement tasks are used routinely as if the methodology is a tried and tested one for studying pragmatics in general, and implicature specifically. In the present paper we reported response accuracy and, for the first time, reaction time evidence, showing that this is not the case. Modifying judgement tasks to include a ternary as opposed to a binary response may be one suitable adaptation, while paradigms that do not involve passing a judgement at all, such as picture selection (Katsos & Bishop 2011, Experiment 3), eye-tracking (Huang & Snedeker 2009) or an action-based task (Pouscoulous et al. 2009), are among many other available solutions .

We should clarify that we do not suggest that all research that has used the binary judgement task to date is deficient. To take a recent example, Skordos and Papafragou (2016) set out to investigate the role of the contextual relevance of the stronger alternative term. They found that 5-year old children's rates of rejection of under-informative utterances were higher when the alternative is relevant for the context than when it is not. This is clear evidence for their hypothesized effect, and the fact that the evidence was demonstrated using a binary rather than any other measure (be it ternary judgement task, picture selection, eye-tracking, or action-based) is largely irrelevant. However, for studies that set out to investigate the age of acquisition of a certain competence or the presence or absence of competence in a certain population (such as in populations with Autism Spectrum Disorders), binary judgement tasks have the potential to substantially underestimate participants' actual competence.

References

Barner, David, Neon Brooks & Alan Bale. 2011. Accessing the unsaid: The role of scalar alternatives in children's pragmatic inference. *Cognition* 118(1). 84–93.

Bott, Lewis & Ira A. Noveck. 2004. Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language* 51(3). 437–457.

Clark, Eve V. 1990. On the pragmatics of contrast. *Journal of Child Language* 17(2). 417–431.

Davies, Catherine & Napoleon Katsos. 2010. Over-informative children: Production/comprehension asymmetry or tolerance to pragmatic violations? *Lingua* 120(8). 1956–1972.

Feeney, Aidan, Susan Scrafton, Amber Duckworth & Simon J. Handley. 2004. The story of some: Everyday pragmatic inference by children and adults. *Canadian Journal of Experimental Psychology* 58(2). 121.

Foppolo, Francesca, Maria Teresa Guasti & Gennaro Chierchia. 2012. Scalar implicatures in child language: Give children a chance. *Language Learning and Development* 8(4). 365–394.

Geurts, Bart. 2010. *Quantity implicature*. Cambridge, UK: Cambridge University Press.

Guasti, Maria Teresa, Gennaro Chierchia, Stephen Crain, Francesca Foppolo, Andrea Gualmini & Luisa Meroni. 2005. Why children and adults sometimes (but not always) compute implicatures. *Language and Cognitive Processes* 20(5). 667–696.

Grice, Herbert Paul. 1975. Logic and conversation. In Peter Cole & Jerry L. Morgan (eds.), *Syntax and semantics*, vol. 3, 41–58. New York: Academic Press (Reprinted in Grice, H. P. (1989). *Studies in the way of words*. Harvard University Press: Cambridge, MA).

Huang, Yi Ting & Jesse Snedeker. 2009. Online interpretation of scalar quantifiers: Insight into the semantics-pragmatics interface. *Cognitive Psychology* 58(3). 376–415.

Katsos, Napoleon. 2014. Scalar implicature. In Danielle Matthews (ed). *Pragmatic Development in First Language Acquisition*, 183–198. Amsterdam: John Benjamins.

Katsos, Napoleon & Dorothy V. Bishop. 2008. Pragmatic Tolerance. Paper presented at the *XI International Congress for the Study of Child Language (IASCL)*, Edinburgh, UK, 28 July–1 August.

Katsos, Napoleon & Dorothy V. Bishop. 2011. Pragmatic tolerance: Implications for the acquisition of informativeness and implicature. *Cognition* 120(1). 67–81.

Katsos, Napoleon & Nafsika Smith. 2010. Pragmatic Tolerance and speaker-comprehender asymmetries, 221–232. In Katie Franich, Kate M. & Lauren L. Keil (eds.), *Proceedings of the 34th Boston University Conference in Language Development*. Cascadilla Press, MA, USA.

Marchena, Ashley de, Inge-Marie Eigsti, Amanda Worek, Kim Emiko Ono & Jesse Snedeker. 2011. Mutual exclusivity in autism spectrum disorders: Testing the pragmatic hypothesis. *Cognition* 119(1). 96–113.

Neys, Wim De & Walter Schaeken. 2007. When people are more logical under cognitive load: Dual task impact on scalar implicature. *Experimental Psychology* 54(2). 128–133.

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

Noveck, Ira A. & Andres Posada. 2003. Characterizing the time course of an implicature: An evoked potentials study. *Brain and Language* 85(2). 203–210.

Noveck, Ira A. & Anne Reboul. 2008. Experimental pragmatics: A Gricean turn in the study of language. *Trends in Cognitive Sciences* 12(11). 425–431.

Papafragou, Anna & Julien Musolino. 2003. Scalar implicatures: Experiments at the semantics-pragmatics interface. *Cognition* 86(3). 253–282.

Papafragou, Anna & Niki Tantalou. 2004. Children's computation of implicatures. *Language Acquisition* 12(1). 71–82.

Pouscoulous, Nausicaa, Ira A. Noveck, Guy Politzer & Anne Bastide. 2007. A developmental investigation of processing costs in implicature production. *Language Acquisition* 14(4). 347–375.

Skordos, Dimitrios & Anna Papafragou. 2016. Children's derivation of scalar implicatures: Alternatives and relevance. *Cognition* 153. 6–18.

Corresponding author's address

Alma Veenstra

Department of Theoretical and Applied Linguistics

English Faculty Building, Room TR 20

University of Cambridge

9 West Road

Cambridge

CB3 9DP

United Kingdom

amv36@cam.ac.uk

Biographical notes

Alma Veenstra is a Research Associate at the Department of Theoretical and Applied Linguistics at the University of Cambridge, UK. She received her doctorate from the Max Planck Institute for Psycholinguistics in Nijmegen, the Netherlands, and held a postdoctoral position at the Université Libre de Bruxelles, Belgium. A psycholinguist by training, her research focuses on language processing and development, with a special interest for pragmatics (the role of Pragmatic Tolerance in the acquisition of implicature) and morpho-syntax (the role of executive functions during the production of subject-verb agreement).

Bart Hollebrandse is a Researcher at the Department of Linguistics at the University of Groningen, the Netherlands. He received his doctorate from the University of Massachusetts at Amherst, USA, and held several international positions. His interests include the theory and acquisition of syntax and semantics, the development of reasoning (Theory of Mind), and the asymmetry in binding between comprehension and production, applying online methods, especially eye-tracking. He also works on recursion and embedding, tense and aspect, focus particles and quantifiers.

Napoleon Katsos is a Reader in Experimental Pragmatics at the Department of Theoretical and Applied Linguistics at the University of Cambridge, UK. He received his doctorate from the University of Cambridge and held several positions here. He is interested in how psychological research can inform theoretical linguistic inquiry and vice versa. His particular focus is in the area of semantics and pragmatics, and in language learning by monolingual and bilingual children as well as children with Specific Language Impairment and/or Autistic Disorders.