Over-informative children: Production/comprehension asymmetry or tolerance to pragmatic violations?

Catherine Davies & Napoleon Katsos

This paper appeared as:

Davies, C. & Katsos, N. (2010) Over-informative children: Production/comprehension asymmetry or tolerance to pragmatic violations? *Lingua* 120 (2010) 1956–1972. doi: 10.1016/j.lingua.2010.02.005

Abstract

It has been proposed that children may fail to comprehend pragmatic aspects of meaning, even though they have little difficulty producing pragmatically appropriate utterances as speakers. A novel account, the pragmatic tolerance hypothesis, predicts that in certain cases children are in fact pragmatically competent both as speakers and as comprehenders and what develops with age is their metalinguistic awareness about accepting or rejecting pragmatically infelicitous utterances as comprehenders. In this paper we focus on pragmatic violations due to over-informativeness and report an investigation into English-speaking 5-year-old children's and adults' production and evaluation of referring expressions. Like adults, children do not over-inform as speakers but unlike adults, child comprehenders do not reject over-informative utterances when given a binary judgment choice. However, when given a magnitude estimation scale which allows for intermediate responses, child comprehenders do penalise over-informative utterances and rate them lower than optimal ones. The findings support the pragmatic tolerance hypothesis. By comparing over-informativeness to under-informativeness and other cases of pragmatic inferencing we propose a set of constraints that predict which pragmatic violations child comprehenders could be tolerant towards.

1. Introduction: over-informing in referential communication

Grice's (1975/1989) second maxim of quantity enjoins interlocutors not to give more information than is necessary for the purposes of the communicative interaction. Take for example a visual display with a single large cup among several other objects, some of which are large but none of which is another cup. The second maxim of quantity predicts that Fred would optimally instruct Bob to pick up this cup by saying 'pick up the cup', and that it would be infelicitous, i.e. over-informative for Fred to do so by saying 'pick up the large cup', since reference to the cup without the adjective would suffice for identifying the referent in this context. If Fred did say 'pick up the large cup', Bob would have to concede that his interlocutor is not obeying the second maxim of quantity. Perhaps he is compelled to

be exceptionally helpful (Clark and Wilkes-Gibbs, 1986), or he has simply not checked that there is only one cup in the display (Pechmann, 1989). On the other hand, if Bob does not concede that Fred is acting against the second maxim of quantity, he would have to infer that Fred wishes to contrast the large cup from a non-large cup, which is nevertheless not present in the display.

Of course, referring expressions (hereafter REs) are optimal or over-/under-informative relative to a communicative purpose. If Bob were to ask Fred 'what is this object?' while pointing to the cup, it would not have been over-informative for Fred to reply 'a large cup', since the purpose of this communicative interaction is to attribute properties rather than to single out the object under discussion. Let us call this the descriptive use of adjectival modification, and conversely, let us call the use of modification for uniquely identifying referents the contrastive use (cf. Karmiloff-Smith's descriptor/determinor functions, 1979:46).

There is a well-established experimental literature on the production of referring expressions in conditions where the communicative purpose is contrastive, i.e. to uniquely identify a referent (e.g. the large cup) out of a range of competitors (e.g. a small cup) and distractors (e.g. a large bowl) (see Deutsch and Pechmann, 1982; Mangold and Pobel, 1988; Pechmann, 1989; Brennan and Clark, 1996; Brown-Schmidt and Tanenhaus, 2006; i.a.). The literature largely documents that in situations with few objects and clearly distinguishing attributes, rates of over-informing are low in production. A smaller pool of research has investigated whether comprehenders reject or penalise over-informative utterances, and the emerging evidence suggests that they indeed do so (Davies and Katsos, 2009, contra Engelhardt et al., 2006). Overall, it seems that adults are Gricean in the sense that they avoid over-informing as speakers and that they penalise over-informativeness as comprehenders[1].

2. The development of informativeness in referential expressions

With regard to child language development, a similar pattern to the adults has been documented in production, whereby children produce few over-informing expressions. In fact, studies have found children to err on the side of under-informativeness in their spontaneous description of object. Dickson (1982) gives a review of the early studies; we provide an overview of some of the key findings in this section.

Informativeness in referring expressions was documented by Sonnenschein (1982), who found that younger children (4–6 years) are more likely to produce under-informative responses than their older peers (7–9 years). Sonnenschein's study suggests that children first go through a period of under-

[1] Throughout the experimental report and discussion, the term *comprehenders* is used to refer to participants acting as off-line raters of utterances, as detailed in the methods sections.

informing, then after around 7 years begin to provide both optimal and over-informative descriptions when they have developed the necessary resources to engage in comparison activity and incorporate the resulting critical features into their REs. The redundant attributes produced by the older child group are claimed to be an easier type of expression to produce than optimal expressions as they do not require full contrastive analysis, so are common in more complex arrays. A similar pattern of redundancy in older children was found by Ford and Olson (1975) who elicited referring expressions from 5- and 7-year-olds in arrays with increasingly large numbers of distractors. The children varied their REs in line with the visual context, articulating the distinctive features within the concurrent array, suggesting that they do engage in comparisons between target items and distractors. Whereas the 5-year-olds gave more minimally contrastive descriptions (e.g. 'the big one' when the target was big and the competitor small), the older children produced more over-informative expressions.

Matthews et al. (2007) note that young children have difficulty producing optimal referring expressions in complex arrays, and are likely to be under-informative; one explanation is that they do not realise that to refer means to describe differences. However, this inadequacy seems to be a matter of performance—while children tend towards under-informativeness in spontaneous referring, they can become more optimal if requests for clarification are made. Matthews et al found that training and feedback result in well adapted expressions, not merely indiscriminately longer ones, replicating earlier work by Whitehurst (1976), in which modelling elicited contrastive (yet over-informative) expressions by children and eradicated under-informative ones. This suggests that as minimally contrastive REs do not widely occur in the naturalistic speech of under-6's, these children are nevertheless able to produce them when trained.

Such studies suggest that children under 6 years of age spontaneously produce under-informative REs, with untrained optimal and over-informative expressions emerging later. If there are developmental changes in the production of informative expressions, are there corresponding changes in informativeness when the child is in the role of comprehender?

Although there is some experimental work on adults' comprehension of over-informative expressions (Mangold and Pobel, 1988; Pechmann, 1989; Maes et al., 2004; Arts, 2004), there is scant research on this phenomenon in development. Sonnenschein (1982) reports a task in which 5- and 9-year-old children had to identify a target group of objects using either optimal or over-informative expressions. The older children were facilitated by over-informing in difficult arrays but younger children were not. Sonnenschein accounts for this with reference to resource-limitation, i.e. retaining a model of the over-informative expression is beyond the memory capacity of the youngest children.

While Sonnenschein's (1982) work addresses the consequences of over-informative expressions in terms of facilitating comprehension in complex arrays, it does not directly address the question of whether in the most simple cases (such as the situation described in section 1), child comprehenders realise that an over-informative expression has deviated from the optimal norm. Yet whether children

are sensitive to violations of the maxim that enjoins interlocutors not to give more information than is necessary for the conversation constitutes a fundamental question in the development of pragmatic competence. This ability presumably underlies successful performance in the production of REs as well as the ability to generate contrastive inferences in comprehension and to classify an interlocutor as cooperative (in terms of quantity of information given) or not. In this paper we will investigate whether indeed 5-year-old English-speaking children are sensitive to the violation triggered by over-informative expressions and what ramifications this may have for theories on the development of communicative competence. We will do so from a speaker's perspective, aiming to corroborate the existing body of research, but also from a comprehender's perspective, where less work has been done.

Since in standard pragmatic theory (Grice, 1975/1989; Hirschberg, 1991; Horn, 1984; i.a.) the ability to detect pragmatic violations due to over-informing is related to the ability to do so for under-informing, we will first look at recent work in the latter area. This will lead us to discuss the Pragmatic Tolerance hypothesis, which is presented in respect to over-informativeness in section 4. The novel experimental findings on over-informativeness from the speaker's and the comprehender's perspective by both adults and children and the discussion are presented in section 6.

3. Scalar implicatures and under-informativeness in comprehension

The production and comprehension of optimally informative referring expressions is mediated by Grice's first and second quantity maxim, enjoining interlocutors not to give too little or too much information respectively. In the recent years, there has been a flourishing experimental literature focusing on whether interlocutors are sensitive to violations of under-informativeness. Such work investigates whether participants reject utterances which would be infelicitous if an implicature mandated by the first maxim of quantity is generated (Feeney et al., 2004; Guasti et al., 2005; Huang and Snedeker, 2009; Hurewitz et al., 2006; Katsos, 2009; Noveck, 2001; Papafragou and Musolino, 2003; Papafragou and Tantalou, 2004; Pouscoulous et al., 2007; i.a.). Most of these studies have focussed on a small subset of quantity-1 implicatures, known as scalar implicatures, which involve utterances with quantifying expressions such as 'some' which can enter into logical scales with other quantifiers, e.g. 'all' and 'most'.

For example, the utterance 'some of the Smurfs went on a boat ride' generates the scalar implicature that the propositions with the other, stronger terms on the scale are not the case, i.e. that it is not the case that most or that all of the Smurfs went on a boat ride. In the typical experimental situation, participants are asked to accept or reject an utterance such as 'some of the Smurfs went on a boat ride' when they have witnessed a scenario where all the Smurfs did so. It is consistently found that adults predominantly reject these utterances, presumably because they have generated the scalar implicature that not all the Smurfs did so. However young children do not reject under-informative utterances, presumably because they have accessed only the semantic meaning of 'some of the Smurfs', namely 'at

least one Smurf' without committing to whether all of the Smurfs went or not. As such, a large number of studies (cited above) document that unlike adults, child comprehenders do not generate implicatures arising out of the first maxim of quantity.

Recently, Katsos and Bishop (submitted for publication) and Katsos and Smith (2010) tested both production and comprehension of under-informative utterances by child and adult participants. Participants watched scenarios unfold on-screen, and then offered binary judgments on an accompanying character's summary of what happened. The critical utterances involved the character offering a semantically true but pragmatically under-informative summary, e.g. utterances containing quantifiers as in the example above, as well as utterances with noun phrases such as ''the dog painted the square'', when in fact the dog had painted both a square and a triangle. Production data were also elicited by asking the participants to answer what happened for themselves. The adults were found to be optimally informative both as speakers and comprehenders, producing all the relevant information when asked to answer the question themselves, and always objecting to the character's under-informative utterances. Moreover, over 90% of the 5-year-old native English-speaking sample was found to be fully informative speakers. However, they were predominantly under-informative comprehenders, rejecting under-informative utterances at rates below 30%. It should be noted that children were 100% able to reject false utterances, documenting that they did not have a problem with rejection in general. On the face of it, these findings corroborate the existing literature on under-informativeness and children's inability to generate quantity implicatures, but add the dimension that this failure in comprehension is in contrast to adult-like production.

Rather than accounting for this discrepancy in terms of a genuine production–comprehension asymmetry, Katsos and Bishop hypothesised that low child performance in comprehension is an artefact of the task with which under-informativeness has been studied. The critical observation was that violations of pragmatics, unlike violations of semantics, do not lead to utterances that are strictly speaking false. As such, children may refrain from rejecting under-informative utterances in a binary choice task, since downright rejection is the response appropriate to falsity. The tendency to differentiate between false and under-informative utterances was in fact found in the adults in an indirect way: while all objections to semantically false utterances were straightforwardly 'no', many under-informative utterances were objected to by accepting that the utterance was true, but highlighting that something important was not mentioned (e.g. adults would reply 'yes, that's right, but the dog painted the triangle as well'). To investigate this apparent task-effect, a follow-up experiment was run using the same story materials and task administration. Children were asked to award the fictional character between one and five strawberries depending on how well he answered the question, rather than to categorically judge whether the utterance was right or wrong. Children awarded a mean of 1.3 strawberries for semantically false utterances in accordance with the universal rejection of these utterances in the previous experiment. They also rated optimal utterances (semantically true and informative) with 4.7, again in accordance with the acceptance of optimal utterances previously. However, they rated under-informative utterances at 3.7, significantly lower than

optimal but higher than semantically false. These findings document that children do in fact penalise utterances with an infelicitous Quantity implicature and as such they are sensitive to the first Gricean maxim of quantity.

In a similar vein, Katsos and Smith (in press) replicated children's early failure with a binary judgment task, but demonstrated that children are in fact informative comprehenders in a sentence-to-picture matching and an act-out task, neither of which involved rejecting pragmatically infelicitous utterances. Thus, Katsos and colleagues conclude that 5-year-old children are informative speakers and they appear to be under-informative comprehenders in sentence-evaluation tasks with a binary judgment. They put forward the pragmatic tolerance hypothesis, whereby while children are indeed pragmatically competent as comprehenders, they differ from adults in that they do not consider pragmatic violations to be grave enough to warrant a rejection of or objection to the offending utterance.

4. Over-informing and the pragmatic tolerance hypothesis

The pragmatic tolerance hypothesis purports to explain the observed speaker/comprehender asymmetry for under-informativeness. However, it is yet to be demonstrated that this account can extend to other cases of pragmatic and communicative development, such as sensitivity to the second quantity maxim, i.e. over-informativeness.

For the case of over-informativeness, let us return to the example where there are two displays: one with a large cup and a distractor which is not a cup (non-contrasting array), and one with a contrast set of two cups, one large and one small (contrasting array). When instructing their interlocutor to pick up the large cup, the prediction of pragmatic tolerance is that like adults, children should not over-inform as speakers: they should refer to it as 'the cup' in the non-contrasting array, and as 'the large cup' only in the contrasting array. With regards to comprehension, in the non-contrasting display, reference to the cup with adjectival modification as 'the large cup', is over-informative and generates a misleading contrastive inference that there exists another non-large cup in the discourse context.

The pragmatic tolerance hypothesis predicts that even though children are sensitive to the second maxim of quantity, they will not predominantly reject over-informative utterances when given a binary judgment, just as they do not consider violations of under-informativeness to warrant a downright rejection even though they are sensitive to the first maxim of quantity. However, when given a graded rating scale rather than a binary judgment, pragmatic tolerance predicts that child comprehenders would penalise over-informative utterances, because they would now have intermediate points on the scale which can be used for utterances that are not optimal but nevertheless better than semantic falsity or syntactic ill-formedness. This would be because children are in fact sensitive to the infelicity caused by over-informativeness.

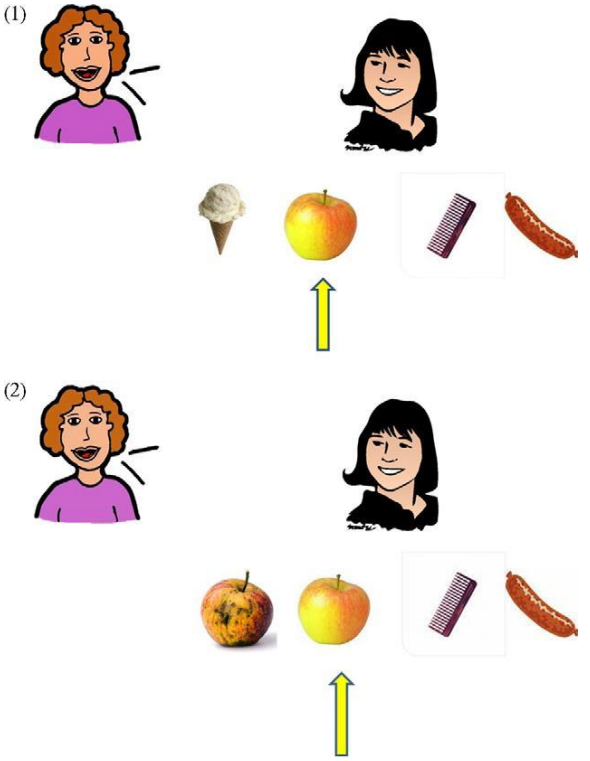
5. Relative penalties for over-informativeness vs. under-informativeness

Having set out the central prediction, let us turn to another prediction of interest. Katsos and colleagues have shown that children are tolerant towards under-informative utterances in the scalar implicature paradigm though not towards semantically false ones. In that paradigm, under-informative utterances are logically true, since it was always indeed the case that the actor had interacted with the object that was mentioned, and reference to the object and the action was appropriate. However, let us turn to the referential communication paradigm and a display where there are two cups, one large and one small. If Fred were to instruct Bob to 'pick up the cup' the utterance would again be under-informative but the overall violation would be graver than in the scalar implicature paradigm, since failure to establish reference precludes evaluating whether the utterance was true or false.

Thus, two different kinds of violations can be triggered by under- and over-informing depending on the experimental paradigm: in the referential communication paradigm where interlocutors have to establish reference to objects to perform certain actions, over-informing ('pick up the large cup') violates a Gricean maxim (quantity-2) but it does not preclude establishing reference and following the instructions. Under-informing ('pick up the cup') on the other hand gives rise to utterances that are not just pragmatically infelicitous but also fail to refer, which leads to the wholesale failure of the directive speech act. However, in the scalar implicatures paradigm where interlocutors are answering wh-questions, neither over- nor under-informing prohibits establishing reference and both violations lead to utterances that are strictly speaking logically true. This observation leads us to a further prediction: in the referential communication paradigm, comprehenders should mostly be tolerant towards over-informativeness. Instead, there should be more severe penalisation in a rating scale and more rejections in a binary judgment for under-informativeness compared to over-informativeness.

In the following sections we present three studies that test the production and comprehension of over- and under-informative utterances by English-speaking children and adults with the aim of testing the predictions spelled out above. Specifically, the production study (experiment 1) measures the incidence of under-, over- and optimal-informativeness using the referential communication paradigm by both adults and children. Studies 2a and 2b document binary judgments and graded ratings respectively of under-, over- and optimal-informative REs by adult and child comprehenders. The goal of these experiments is to test the two predictions of the pragmatic tolerance hypothesis: (a) whether child comprehenders indeed fail to show sensitivity to over-informativeness when making binary judgments of pragmatic felicity, even though from the speaker's perspective they are as sensitive to over-informing as adults are, and (b) to test whether children show sensitivity to violations of over-informativeness when a graded rating scale is available, in line with previous work on the detection of violations of under-informing. A third prediction, which can be accommodated by the pragmatic tolerance hypothesis but is not central to it, is that violations of under-informing in this experimental paradigm

will consistently be judged as worse than violations of over-informing.



Figs. 1 and 2. Experiment 1: example arrays for adults. (1) 1-Referent display; (2) 2-referent display.

6. Experiment 1: production study

6.1. Participants

The child population were 24 monolingual native speakers of English, with a mean age of 5;5 years ranging between 4;10 and 6;5. All were recruited from three primary schools in Cambridge, UK. The same child sample was used for all three experiments in the current paper. 24 undergraduate and graduate students of various disciplines from the University of Cambridge were tested, all of whom were native English-speaking adults.[2]

6.2. Experimental design

In this experiment we measured levels of informativeness in production by adults and 5-year-old children. A computer-based sentence-production task was created by displaying pictures of everyday objects taken from Google images on Microsoft PowerPoint software. Participants were presented with the stimuli and were asked to instruct a fictional character to pick up an object. We manipulated

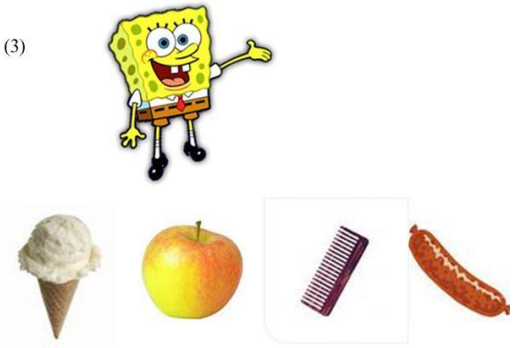
[2] Due to a coding error, the age-range of the adult participants is not available, but the typical mean age for participants from our subject pool is around 24 years.

presence or absence of contrast set (another object of the same type which is contrasted along some attribute, e.g. one fresh and one mouldy apple) in the visual array, creating 1-referent and 2-referent displays (see Figs. 1–4 for illustrations). The task was administered by a single experimenter and the participants' responses were voice-recorded and later analysed. The method varied slightly between the two age groups since we wanted to make the version for 5-year-olds as child-friendly as possible. Owing to these slight differences, the procedures are explained separately below.

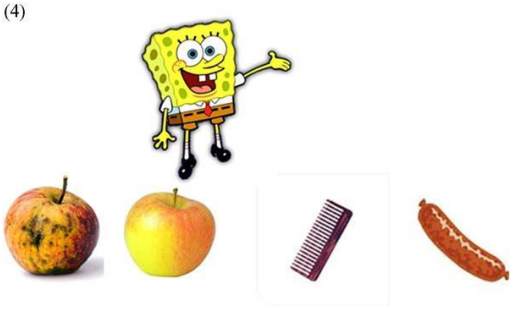
6.2.1. Adult procedure

A static laptop display showed two characters on a computer screen, one of whom had four items in her vicinity on the right-hand side of the screen. See Figs. 1 and 2 for example displays. One of these items was highlighted with an arrow, and participants were asked to state the best way for the character on the left to ask for the cued object. There were two practice items: one with a 1-referent display and one with a 2-referent display. The production study took less than 5 min to administer, and formed part of a longer experimental session, involving additional experiments for other research projects.

(3)



(4)



Figs. 3 and 4. Experiment 1: example arrays for children (on-screen display). The arrow indicating the target object to be referred to was included in the booklet display, which only the children could see. (3) 1-Referent display; (4) 2-referent display.

6.2.2. Child procedure

On a static laptop display children were introduced to the cartoon character SpongeBob SquarePants, who appeared on-screen accompanied by four objects. See Figs. 3 and 4 for example displays. The experimenter then explained that the participant should ask for one of SpongeBob's objects. To convey which object they should ask for, the participant was given an identical array of objects as the ones that SpongeBob SquarePants had, printed in a booklet which SpongeBob could not see. The objects to ask for were cued by an arrow on the relevant page of the booklet. This was done so that it would be

evident to participants that SpongeBob would not know which object to give them unless they describe it appropriately. Participants were instructed to look at all of the objects on the booklet page carefully before asking SpongeBob to give them the right thing. There were two practice items: one with a 1-referent and one with a 2-referent display. The whole experiment took around 5 min to complete.

As the same children took part in experiments 1, 2a and 2b, the production experiment always preceded the two comprehension experiments to avoid influence from over- or under-informing speakers in the comprehension studies. The tasks were also separated by an unrelated experimental task for a separate research project. Together with a sentence-repetition task from the NEPSY battery (Korkman et al., 1999), this formed an experimental session that lasted between 30 and 40 min, conducted in a quiet area in the children's school. The sentence-repetition task was administered in order to identify participants that fell below the norms for their age group and to establish whether participating children were typical for their age. One of the children in the original sample performed below the norms and so was replaced with a new participant. All of the participants in the final sample of 24 performed within 1.5 standard deviations from the age-appropriate mean.

6.2.3. Materials

Descriptions of the target object presented in isolation (without any other objects in the display) had been previously elicited from an independent group of adult native speakers of English (n = 29). Only items which were referred to without using adjectival or other modification in more than 80% of cases in this pretest were added to the stimulus lists for use in the studies reported here. This was done to control for the possibility that a certain attribute is mentioned simply because it is particularly salient in a given stimulus (i.e. a certain shoe might have looked so old that participants would describe it as an 'old shoe' even when no other shoes are in the display). This would interfere with the purposes of the experiment where we aim to study whether participants produce adjectives for contrastive purposes only.

For the adults, 40 items were created, each in two versions, the 1-referent display (no contrast set), and the 2-referent display (contrast set). These 40 items were split into two groups of 20. Half the participants saw the first group of items in the 1-referent display and the second group of items in the 2-referent display; while the other half participants saw the item groups the other way round. See Appendix A for the list of items. The order of items within each group was pseudo-randomised.

For the children, there were 24 items in total (which were a subset of the adults' 40 items) in order to make the experiment shorter. Each item was created in two versions, the 1-referent display (no contrast set), and the 2-referent display (contrast set). The items were split into two groups and presented to participants in a similar fashion as to the adults.

6.3. Coding the responses

All responses were recorded, transcribed and classified as under-, optimally- or over-informative

depending on the visual display. For example, in the array depicted in Fig. 1, an optimally informative expression would be 'pass me the apple', and an over-informative expression would be 'pass me the fresh apple'. 'Pass me the apple' would be under-informative in Fig. 2, with e.g. 'pass me the fresh apple' qualifying as optimal and 'pass me the good, fresh apple' as over-informative. On very few occasions, child participants used a reduced form of the noun (e.g. 'pass me that'), or pointed at the entity on the screen without saying its name. In such cases, participants were reminded that SpongeBob could not understand which object they wanted and that they should try again. When prompted, all participants then referred using a noun phrase.

6.3.1. Results: adults

6.3.1.1. Quantitative analysis. The proportions of under-, optimal, and over-informative referring expressions produced by adults for 1-referent and 2-referent arrays are presented in Fig. 5. With regards to the 1-referent displays, where only two types of output are possible (optimal and over-informative), a Wilcoxon's signed rank test revealed that participants were predominantly optimal (Z = 4.33, p < .001). With regards to the 2-referent displays, where under-, optimal-, and over-informative expressions were elicited, a Friedman test revealed a significant difference between condition means (x^2(2) = 28.74, p < .001). Further pair-wise comparisons by means of Wilcoxon's signed rank tests revealed that participants were predominantly optimal rather than over-informative in the 1-referent condition (Z = 4.33, p < .001), and predominantly optimal rather than under- or over-informative (Z = 2.30 and Z = 4.12 respectively, both ps < .05) in the 2-referent condition. In the latter condition they produced more under-informative compared to over-informative utterances (Z = 2.79, p < .01). When comparing the 1- and 2- referent condition means using a Wilcoxon's test, the rates of over-informativeness remain stable (Z = 1.01, p > 0.3, n.s.), i.e. the adults' rates did not differ as a function of the display type.

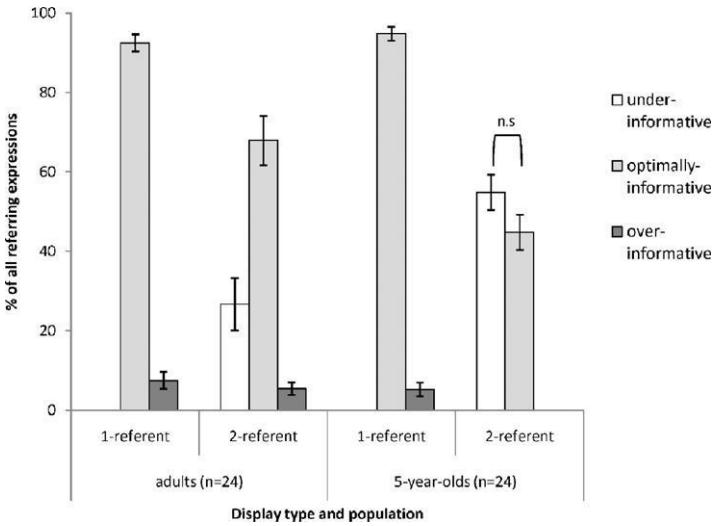


Fig. 5. Experiment 1: proportions of adults' and 5-year-old's under-, optimally-, and over-informative referring expressions (mean rates shown as % of all expressions by display type). All within-participant comparisons are significantly different (p < .001 – p < .05), unless otherwise marked.

Thus, when encountering displays without a contrast set, adult speakers were overwhelmingly optimal. When referring to targets which formed part of a contrast set, speakers were also largely optimal in their referring strategies, though relatively high rates of under-informativeness were also documented. The rates of over-informativeness are the same in the 1- and 2-referent displays, and in both cases they are low.

6.3.1.2. Qualitative analysis. The 56 tokens of over-informative expressions that were elicited from the total of 960 expressions by adults were also analysed qualitatively. Of these, the attribute most frequently provided redundantly was colour (48%). Type adjectives such as handheld (phone) comprised 14% of over-informing expressions and the contents of items such as glass of water 13%. Other attributes provided in less than 5% of cases were (in descending order) size, age, pattern, shape, state (e.g. empty), material, speaker attitude (e.g. nice) and location. The results regarding the use of colour adjectives accord with previous work which suggests that colour attributes are often used gratuitously rather than discriminatively (e.g. Mangold and Pobel, 1988), underscored by Sedivy (2002) who found that unpredictable colour modifiers are frequently encoded even in default descriptions (where items are referred to in isolation, without competitors or distractors).

6.3.2. Results: children

6.3.2.1. Quantitative analysis. Two tokens were not comprehensible in the recorded file and were excluded from the final sample. The proportions of under-, optimal, and over-informative expressions for the remaining REs by 5-year-olds for 1-referent and 2-referent arrays are presented in Fig. 5. With regard to the 1-referent displays, where only two types of output are possible (optimal- and over-informative), a Wilcoxon's signed rank test revealed that children were predominantly optimal ($Z = 4.42$, $p < .001$). With regards to the 2-referent displays, children were optimal and under-informative in equal rates ($Z = 1.09$, $p > .2$, n.s.). However, there were no instances of over-informing at all.

Thus, when encountering displays without a contrast set children were overwhelmingly optimal, just like the adults. When referring to targets which formed part of a contrast set, children were either under-informative or optimal. Just like the adults, children over-informed at very low rates overall. A Mann–Whitney U-test run to compare the rates of under-informing in the 2-referent display across groups, reveals that children under-informed significantly more than adults ($Z = 3.72$, $p < .001$). Children's high rates of under-informing may be due to the same lack of comparison activity between the target and distractors as documented in the previous studies outlined in section 1.

6.3.2.2. Qualitative analysis. The 15 instances of over-informative reference that were elicited from the 574 referring expressions by children were also analysed qualitatively. Of these, the attributes most frequently provided redundantly was colour and size (27% each). Other attributes appearing just once

or twice in the entire data set were pattern, speaker attitude (e.g. smelly), idiosyncratic words (e.g. canny pan) and orientation. The results regarding the colour over-modification accord with both the adult data and previous work, as discussed above.

7. Discussion: experiment 1

In experiment 1 we found that neither children nor adults over-inform. Also, both groups are optimally informative when there is only one instance of a certain entity in an array. However, there are non-negligible rates of under-informativeness in 2-referent displays, 27% and 55% for adults and children respectively. Comparisons showed that children under-inform significantly more than adults. One issue to clarify is whether it is possible that the reason why children do not over-inform is different than the reason why adults do not do so. Specifically, is it possible that the low rates of over-informing in 1-referent displays in children are due to the fact that they generally do not produce adjective–noun constructions rather than that they are sensitive to the second Gricean maxim of quantity? It is unlikely that this is the case, because children do produce an adjective–noun construction in 45% of utterances in the 2-referent display, where the adjective–noun construction is the optimal response, and only in 5% of utterances in the 1-referent display, where the adjective would be over-informative. This indicates that the low rates of adjectival modification in the 1-referent display are due to sensitivity to Gricean maxims and avoidance of over-informing rather than inability to produce adjectival constructions.

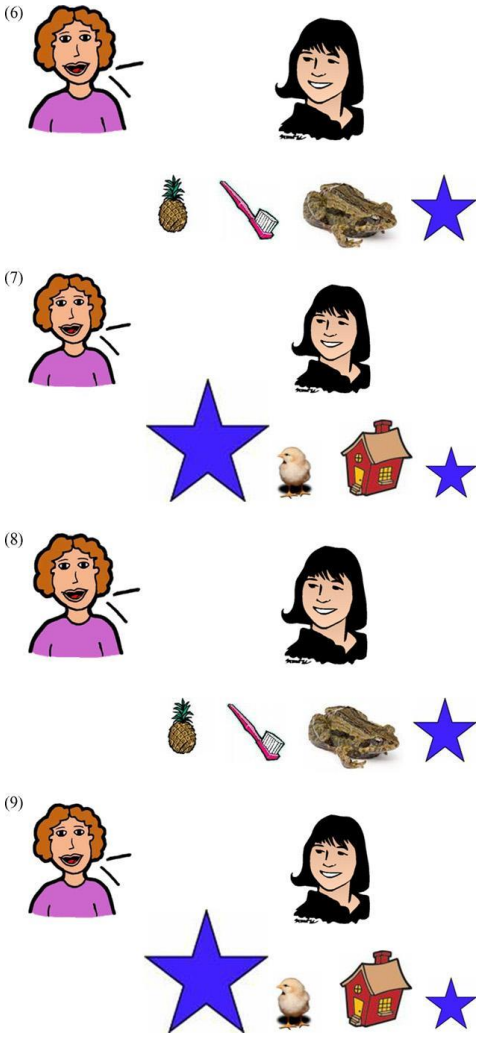
These results from the production of over-informing expressions will be used as a benchmark for experiments 2a and 2b to establish whether there is a speaker/comprehender asymmetry with regards to the acquisition of informativeness.

8. Experiments 2a and 2b: comprehension studies

8.1. Design and participants

In two further experiments we investigated whether children and adults detect violations of informativeness as comprehenders and how they evaluate them. Participants were tested using a computer-based sentence-evaluation task by combining Clipart Pictures with pre-recorded utterances on Microsoft PowerPoint software. The task was again administered by a single experimenter. We used a 2 2 within-subjects design, manipulating presence of a contrast set in the visual array, and presence of adjectival modification in the heard utterance, creating four conditions: over-informative (no contrast set with modification), under-informative (contrast set, no modification) (henceforth Over- and Under-conditions respectively), and two optimal conditions optimal-1 (no contrast set, no modification), and optimal-2 (contrast set with modification) illustrated in Figs. 6–9. In the comparisons to follow, the corresponding optimal condition for Over-utterances is optimal-1 which has the identical visual display

(one object of the critical type) but contrasts with regards to absence of modification. The corresponding optimal condition for Under-utterances is optimal-2 which shares the same visual display (two objects of the critical type) but contrasts with regards to the presence of modification.



Figs.6–9. Experiments 2a and 2b: example arrays for adults.

(6) over-informative: 'pass me the small star';
(7) under-informative: 'pass me the star';
(8) optimal-1: 'pass me the star';
(9) optimal-2: 'pass me the small star'.

Comprehension was tested in two ways: a binary judgment task and a magnitude estimation (ME) scale. In the child sample, the same participants took part in both tasks, and were the same participants who had taken part in the production study (n = 24). Each task was set up separately, and each participant was randomly allocated to a binary judgment first-magnitude estimation second order, or vice versa. For the adults, different participants took part in each comprehension task, and were different participants from those who had taken part in the production task. Twelve native English-speaking undergraduate and graduate students of various disciplines from Cambridge University were tested on the binary judgment task (mean age 24 years old, ranging from 20 to 35 years). Twenty-four participants with the same profile as the binary judgment group took part in the ME task (mean age 24 years old, ranging from 21 to 33 years).

We will report the common sections for each experiment, and then the binary judgment methods and findings for children and adults as experiment 2a and the magnitude estimation ratings for children and adults as experiment 2b. Again, the methods differed slightly between the adults and the children, so we will explain the procedures separately below.

8.1.1. Procedure and materials

The adults made their magnitude estimation ratings in a session lasting around 20 min. There were 60 items in total: 40 critical items with 10 in each critical condition (under-, over-, optimal-1 and optimal-2) plus 20 syntactically ungrammatical items, e.g. 'pass me the cup plastic'. These control items were included to ensure that participants were able to reject and penalise straightforwardly inappropriate utterances (see Appendix B for a list of all items). Every target item appeared in only one of the four conditions between participants, i.e. a Latin square design was used to counterbalance any item effects. The items were presented to participants in either one of two pseudo-randomised orders.

For the children, there were 36 items in total (which were a subset of the items for the adults) to make the experiment shorter; for 18 of these items the children were asked to make a binary judgment and for the other 18 they were asked to rate how well the on-screen character asked for the item using a magnitude estimation scale. Of each eighteen items, twelve critical items were allocated to the four critical conditions. The remaining six items were syntactically ungrammatical utterances. The two tasks (described in detail below) were run consecutively, with equal numbers of participants allocated to each of the two orders of presentation. As a benchmark for successful performance on these tasks, any child that did not reject or penalise the syntactically ungrammatical utterances at rates higher than the mean of the optimal utterances would be excluded from the task on the grounds that they did not comprehend the requirements of the task.

There were four versions of the experiment, comprising the set of target items rotated around each of the four critical conditions in a Latin Square design. No critical target item was seen more than once by any participant, although the syntactically infelicitous items were the same across versions. Each version of the entire set of items was presented in one of two randomised blocks, giving eight versions of the stimuli overall.

9. Experiment 2a: comprehension with a binary judgment task

A static laptop display showed two characters, one of whom had four items in her vicinity (see Figs. 6–9 for example trials). Adult participants heard pre-recorded sound clips of one character asking the other to 'pass me the [referential expression]'. The participants were asked to make a binary judgment as to whether the instruction they heard was natural or unnatural.

For the children, a static laptop display showed two characters, Mickey and Donald, with four objects in Donald's half of the screen. At the beginning of the experiment the same participants as in experiment 1 were introduced to the on-screen characters. The experimenter explained that Donald has four objects on his side of the screen and that Mickey wants one of them, so he asks Donald to pass him the desired object. The same pre-recorded instructions as used with the adults with the form 'pass me the X' were played automatically when each new item began. The target object in each trial was highlighted with an arrow, and the children were told that Donald could not see it. The experimenter elaborated that sometimes Mickey asks in a good, sensible way, and sometimes in a bad, silly way, and that the participant should say whether Mickey asked Donald for the object in a good or a bad way.

9.1. Predictions

Recall from section 4 that the pragmatic tolerance hypothesis predicts that children will not reject violations of over-informativeness using a binary judgment on the grounds that they are not grave enough to warrant a rejection. As such, the account predicts a speaker/comprehender asymmetry for over-informativeness.

9.2. Results

9.2.1. Adults

The results of the binary judgment means by the group of 12 adults are presented in Fig. 10. NB the binary data in the graph have been plotted to show acceptances rather than rejections to allow simpler comparison with the positive magnitude estimation ratings presented in section 11. A Friedman test revealed a significant difference between critical conditions ($x^2(3) = 17.59$, $p < .001$). Further pair-wise comparisons by means of Wilcoxon's signed rank tests reveal that over-informative utterances receive more rejections than their optimal-1 counterparts ($Z = 2.53$, $p < .05$), and that likewise, under-informative utterances receive more rejections than their optimal-2 counterparts ($Z = 2.57$, $p < .05$). When comparing by infelicity type, results show that under-informative utterances received more rejections than their over-informative counterparts ($Z = 2.10$, $p < .05$). This is in line with the discussion in section 5 where the two violations give rise to different kind of communicative failure in the referential communication paradigm.

9.2.2. Results: 5-year-olds

Every child met the benchmark for grasping the task, i.e. to reject the ungrammatical utterances at rates higher than the optimal ones, and was therefore kept in the sample of data to be analysed. The results of the binary judgment means of the 24 children are presented in Fig. 10, plotted to show acceptances. A Friedman test revealed significant differences between critical conditions ($x^2(3) = 8.97$, $p > .05$). Further pair-wise comparisons by means of Wilcoxon's signed rank tests reveal that over-informative utterances are not treated differently from than their optimal-1 counterparts ($Z = 1.51$, $p > .1$, n.s.). In contrast, under-informative utterances are rejected more than their optimal-2 counterparts ($Z = 2.36$, $p < .05$).When comparing by infelicity type, under-informative utterances numerically received

more rejections than their over-informative counterparts (means of 0.25 vs. 0.11 respectively), but this difference did not reach significance ($Z = 1.85$, $p > .05$, n.s.).

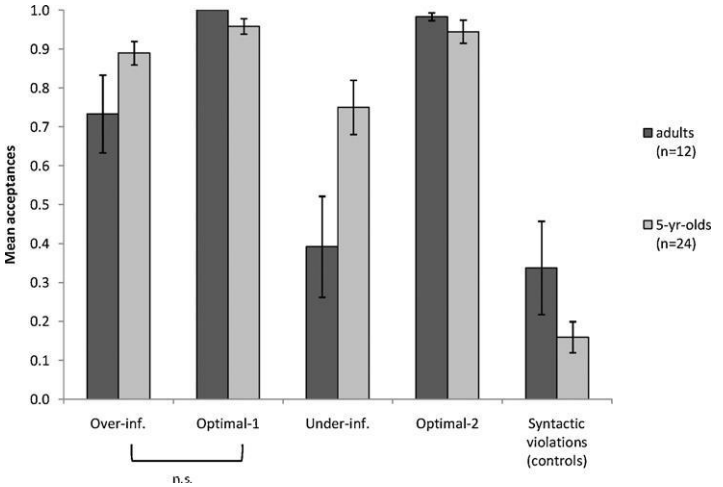


Fig. 10. Experiment 2a: binary judgments: mean acceptances by adults and 5-year-olds. All within-participant comparisons are significantly different ($p < .05$), unless otherwise marked.

10. Discussion: experiments 1 and 2a

The results from experiments 1 and 2a show that adults avoid over- and under-informing as speakers, and that they reject both over- and under-informative utterances more than their felicitous counterparts when making a binary judgment. However, the children do not reject over-informative utterances at rates higher than their optimal counterpart, even though as speakers they avoided over-informing. These results document a speaker/comprehender asymmetry in children with regards to avoiding and rejecting over-informativeness, as predicted by the pragmatic tolerance account.

The wholesale acceptance of over-informative utterances by children is in line with pragmatic tolerance. Nevertheless, for this account to be directly validated it must be further demonstrated that in a non-binary judgment task, the very same children that took part in experiment 2a will be able to detect and penalise violations of over-informativeness. If this is borne out, it will document that children of this age group are not too young to master quantity-2 and that their performance in 2a is not driven by lack of pragmatic competence. This is tested in experiment 2b which instead of a binary judgment asks participants to rate their response on a scale.

11. Experiment 2b: comprehension with a magnitude estimation rating

11.1. Method and procedure

11.1.1. Adults

The adult population is the same as the one that took part in experiment 2a. A static laptop display

showed two characters, one of whom had four items in her vicinity (see Figs. 6–9 for example trials). Participants heard pre-recorded sound clips of one character asking the other to 'pass me the [referential expression]' and then rated how natural the instruction was using a magnitude estimation (ME) scale (Bard et al., 1996), with a syntactic violation used as the modulus. Participants noted down their ratings with paper and pen. A 5-min training session on ME judgments was administered by the experimenter with all participants before the experiment began. ME is a technique which allows participants to make fine-grained acceptability judgments by designing their own scale which has no upper or lower limit. All judgments are made relative to a modulus, which is an item estimated to be at the mid-range of acceptability. ME ratings are particularly recommended for investigating participants' sensitivity to moderately mild violations, i.e. violations that are not overriding a core grammatical rule.

11.1.2. 5-Year-olds

A static laptop display showed two characters, Mickey and Donald, with four objects in Donald's half of the screen. At the beginning of the experiment the same participants as in experiment 1 (n = 24) were introduced to the on-screen Mickey and Donald. The experimenter explained that Donald has four objects on his side of the screen and that Mickey wants one of them, so he asks Donald to pass him the desired object. The same pre-recorded commands as used with the adults took the form 'pass me the X' and played automatically when a new item began. The target object in each trial was highlighted with an arrow, and the children were told that Donald could not see it. The experimenter elaborated that sometimes Mickey asks in a good, sensible way, and sometimes in a bad, silly way, and that there was a big basket of strawberries to be used as rewards (depicted as a colour-printout in front of the children). The participants were told that they should give Mickey as many strawberries as he deserves (with no upper limit) for the way he has asked for the object, but that they should always give at least one strawberry since he had tried hard (in line with ME requiring a positive numerical rating). Two example items and practice ratings were conducted. The ME rating given in the practice trial was used as a modulus and the participants were verbally reminded of this rating throughout the experiment.

11.2. Results

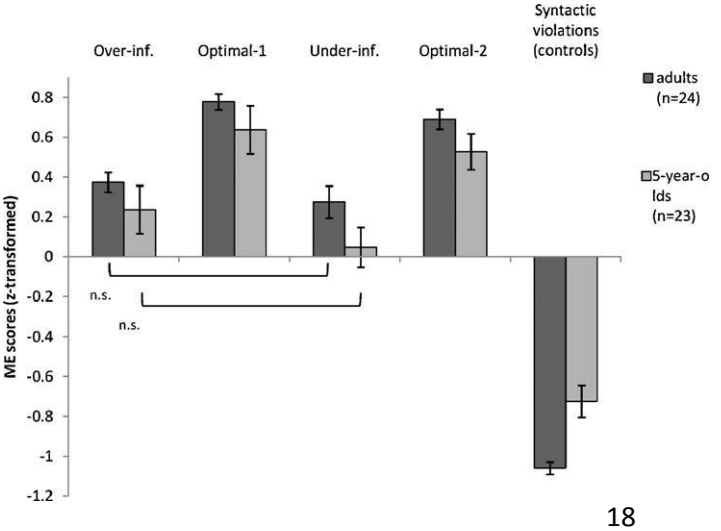


Fig. 11. Experiment 2b: magnitude estimation (z-transformed): mean ratings by adults and 5-year-olds. All within-participant comparisons are significantly different ($p < .001$–$p < .05$), unless otherwise marked.

11.2.1. Adults

The ME ratings were normalized (by means of z-scores) as is standard practice (cf. Bard et al., 1996) and means from the normalized ratings are presented in Fig. 11. A 2 x 2 repeated measures ANOVA for the four critical conditions reveals no main effect of contrast ($F(1, 23) = 3.63$, $p > .06$, n.s.) and no main effect of modification ($F(1, 23) = .01$, $p > .9$, n.s.). Instead, there was a significant interaction between contrast and modification ($F(1, 23) = 21.64$, $p < .001$). Further pair-wise planned comparisons by means of t-tests show that over-informative and under-informative utterances were both rated lower than their corresponding optimal utterances, i.e. optimal-1 and optimal-2 ($t(23) = 4.90$, $p < .001$; $t(23) = 3.47$, $p < .005$; respectively). Over-informative utterances were numerically rated higher than under-informative utterances, although this difference did not reach statistical significance ($t(23) = 1.08$, $p > .2$, n.s.).

11.2.2. 5-Year-olds

The child population is the same that took part in experiments 1 and 2a. One child rated the ungrammatical utterances higher than the mean of the optimal ones, even though in the binary judgment task they rejected the ungrammatical utterances more than the optimal ones. This suggests that they did not grasp the magnitude estimation task, and so their data were removed from the final sample. The normalized data (by means of z-scores) from the magnitude estimation ratings from the remaining 23 children are presented in Fig. 11. A 2 x 2 repeated measures ANOVA for the four critical conditions reveals no main effect of contrast ($F(1, 22 = 1.73$, $p > .2$, n.s.), and no main effect of modification ($F(1, 22) = .11$, $p > .7$, n.s.). There was a highly significant interaction between contrast set and modification ($F(1, 22) = 13.51$, $p < .001$), documenting that 5-year-olds were sensitive to both modified and unmodified utterances in both felicitous and infelicitous contexts. Further pair-wise planned comparisons by means of t-tests show that over-informative utterances were rated lower than their corresponding optimal-1 utterances ($t(22) = 2.29$, $p < .05$) and so too were under-informative utterances compared to optimal-2 ($t(22) = 2.93$, $p < .005$). There was a tendency for the over-informative utterances to be rated higher than the under-informative utterances although this difference did not reach statistical significance ($t(22) = 1.01$, $p > .3$, n.s.).

The results from experiment 2b show that adults penalise both over- and under-informative utterances more than their felicitous counterparts. Numerically, adult comprehenders are also less severe towards over-informative than under-informative utterances. A very similar picture is obtained with the 5-year-old children. These results indicate that 5-year-old children are in fact perfectly sensitive to violations of informativeness. Specifically, corroborating the findings of Katsos and Bishop (submitted for publication) and Katsos and Smith (in press), it is found that children are sensitive to violations of the first maxim of quantity, which give rise to under-informative utterances. More critically, children are indeed sensitive to violations of the second Gricean maxim of quantity which are triggered by over-informative utterances.

12. General discussion

First, a theory-neutral finding is that in both experiments 2a and 2b, ungrammatical utterances are rejected and penalised at higher rates than any of the two pragmatically infelicitous over- and under-informative utterances. This holds for the 5-year-old children, but more crucially it also holds for the fully competent adult group as well. The fact that pragmatic violations are treated differently than violations of syntactic well-formedness documents that the constructions under investigation are subject to psycholinguistically as well as linguistically distinct types of rules.

With regard to the critical question we set out to address, whether children are sensitive to the second maxim of quantity as speakers and comprehenders the picture that emerges from experiments 1, 2a and 2b together is that like adults, children avoid over-informing as speakers (experiment 1), but unlike adults they do not reject over-informative utterances in comprehension when given a binary judgment (experiment 2a). However, this was not the case because they were too young to master the second Gricean maxim of quantity altogether: when given a magnitude estimation scale in experiment 2b, the same children penalised over-informative utterances less than semantically false ones but more than optimal ones. Our findings are straightforwardly accommodated within the pragmatic tolerance hypothesis which predicts that young children do not treat pragmatic violations as grave enough to warrant a rejection when the only option is a binary acceptance or rejection. However, when given a range of responses where children can select intermediate responses, they clearly show a penalisation for over-informativeness.

Furthermore, it was found that both in the binary and the magnitude estimation task, and both in adults and children, under-informative utterances are rejected more and penalised more severely than over-informative utterances in the paradigm that we used. This is in line with the prediction that pragmatic violations that preclude establishing reference are treated as more severe than pragmatic violations that result to otherwise felicitous (in the sense of establishing reference) and logically true utterances. This finding contributes to a list of constraints on when children should be tolerant towards pragmatic infelicity: Recall, that using the scalar implicature paradigm, Katsos and colleagues document that children are tolerant towards violations of under-informativeness which do not lead to a breakdown of reference assignment, but that this is not the case when using the referential communication paradigm, where under-informativeness leads to reference assignment failure. Children are also found to display tolerance to over-informativeness in referential communication, where the infelicity does not impede reference establishment. Moreover, it is found that children are not tolerant whatsoever towards semantic falsity (Katsos and Bishop, submitted for publication; Katsos and Smith, in press), and syntactic ungrammaticality (present study). As such, this account predicts tolerance towards violations that are exclusively pragmatic in nature, and moreover only in cases where the pragmatic violation does not preclude success with fundamental aspects of the speech act, such as reference assignment.

In our review of the literature on the development of informativeness in referential communication in section 2, two patterns emerged: spontaneous under-informativeness in under-6's (Sonnenschein, 1982; Matthews et al., 2007), and increased over-informativeness after the seventh year (Sonnenschein, 1982; Ford and Olson, 1975). In our sample, 5-year-olds under-specified the target in the presence of a contrast set (2-referent condition) in 55% of trials (numerically but not statistically more than the incidence of optimal REs) which corroborates the previous findings. Moreover, they did not produce over-informative REs in the presence of a contrast set. It is possible that redundancy of this type may indeed emerge later on in development, to level out at around 6% in the adult population.

In comprehension, previous work by Sonnenschein (1982) reveals that younger children are hindered by over-informative REs in complex displays. Although Sonnenschein's approach does not share the same theoretical assumptions as the current study, it may have been the same emerging sensitivity to non-default, over-informative expressions which we elicited in our ratings studies which impeded the 5-year-olds in identifying targets in Sonnenschein's study. However, the differences in the materials used and task performed hinder a full comparison with the present study.

13. Conclusions

In their aims to develop and test the emerging pragmatic tolerance hypothesis, the results of the present study and its forerunners suggest that the production/comprehension asymmetry apparent on administering binary judgment tasks with children is an experimental artefact. However, prima facie, children's adult-like adherence to the second quantity maxim in production and their seeming insensitivity to its violation in comprehension might lead one to turn to an account such as Interface Asymmetry (Reinhart, 2004, 2006). Reinhart's account of phenomena at the interface of the computational system and pragmatics indeed predicts no delays to arise in production while comprehension is said to be compromised due to working memory limitations concerning reference-set computation (see also Hendriks and Koster, this issue for an overview). Applied to the phenomenon with which we are concerned, namely the sensitivity to and comprehension of over-informative utterances, the interface asymmetry account predicts that even if children are able to generate the reference-set of under- and optimally informative referring expressions as comparisons to the heard over-informative counterpart, they would reject the over-informative expressions at chance levels by resorting to guessing or fixed-response strategies. However, against the predictions of the interface asymmetry account, our child participants wholesale accepted over-informative utterances rather than rejecting them at rates of around 50%. Moreover, this was not the case because they were too young to master the second Gricean maxim of quantity altogether: when given a magnitude estimation scale in experiment 2b, the same children did penalise over-informative utterances. Thus, interface asymmetry cannot account for the results presented in our studies.

Finally, we will further briefly touch upon two issues: The first is methodological and it pertains to the

tasks that are not susceptible to pragmatic tolerance and therefore are most recommended for revealing children's actual competence. It should be clarified that the juxtaposition of a binary judgment task with a magnitude estimation scale in the present does not in any way mean that a magnitude estimation scale is necessarily the preferred non-binary judgment task to be used. Act-out tasks, sentence-to-picture matching tasks (such as the ones reported in Katsos and Smith) and even more indirect tasks such as eye-tracking during comprehension all have in common the fact that they do not coerce children's competence to manifest into a binary judgment. Magnitude estimation scale ratings include a metalinguistic component (since participants do not just have to detect a violation but also to evaluate how severe it is), so, if anything, it is preferred less than the other non-metalinguistic tasks.

The remaining (and central) question for pragmatic tolerance concerns what exactly develops with time, turning tolerant children into intolerant adults? Since judgment tasks have a metalinguistic component, and force a binary choice, a key skill that develops with age ought to be the metalinguistic awareness that even slight violations of meaning (pragmatics) ought to be rejected or objected to in a similar fashion as violations of truth (semantics) and well-formedness (syntax). Alternatively, it could be that children accept pragmatic infelicity in binary choice because they cannot actively introduce a third category, distinct from false and optimal. Recall that the adults in Katsos and Bishop's binary judgment study often indicated that under-informative utterances where right but also somehow lacking (making comments such as 'that's right, but . . .' or that was 'half right, half wrong'). It could be that children are following the same tendency which prompts adults to accept the pragmatically infelicitous utterance but they lack the adult ability to conceptualise how to further express that something is missing. Clearly, more conceptual and empirical work is required to address this issue.

Acknowledgements

We are grateful to the headteachers, staff, parents and students of the Sancton Wood Primary School, Cambridge International School, and Parkside Primary School, all of Cambridge, UK, for participating in this research. The first author is supported by a Domestic Research Scholarship from the University of Cambridge.

Appendix A. Sample items for experiment 1 (adult version)

Item	Target	Competitor (2-referent condition only)			Optimal referring expression
1-ref01	thin nail	fork	apple	cat	the nail
1-ref02	tall boot	carrot	tree	microphone	the boot
1-ref03	new phone	bike	compass	feather	the phone
1-ref04	old computer	watch	guitar	mirror	the computer
1-ref05	intact egg	onion	camera	hairdryer	the egg
1-ref06	closed umbrella	onion	ice-cream	stool	the umbrella
1-ref07	open book	hamburger	fish	teddy	the book
1-ref08	unsliced bread	toothbrush	toad	pineapple	the bread
1-ref09	unlit cigarette	tiger	plane	soap	the cigarette
1-ref10	modern rug	banana	crab	toaster	the rug
1-ref11	banana	cube	fork	watch	the banana
1-ref12	comb	onion	hamburger	toothbrush	the comb
1-ref13	hammer	apple	tiger	tree	the hammer
1-ref14	pear	camera	ice-cream	guitar	the pear
1-ref15	sausage	toad	plane	crab	the sausage
1-ref16	boat	cat	microphone	feather	the boat
1-ref17	drum	stool	hairdryer	teddy	the drum
1-ref18	corkscrew	toaster	duck	soap	the corkscrew
1-ref19	flower	razor	anchor	drum	the flower
1-ref20	onion	cup	fridge	vase	the onion
2-ref01	closed bag	open bag	strawberry	zebra	the closed bag
2-ref02	old newspaper	recent newspaper	cherry	dog	the old newspaper
2-ref03	stripy cup	spotty cup	tv	cucumber	the stripy cup
2-ref04	new phone	old phone	hat	flower	the modern/cordless phone
2-ref05	sleeping baby	feeding baby	butterfly	tap	the sleeping baby
2-ref06	full glass	empty glass	iron	broom	the full glass
2-ref07	long skirt	short skirt	football	toilet	the long skirt
2-ref08	short sock	long sock	spoon	lobster	the short sock
2-ref09	tall vase	short vase	key	bus	the tall vase
2-ref10	small star	big star	house	chick	the small star
2-ref11	glass mug	china mug	duck	corkscrew	the glass mug
2-ref12	adult penguin	baby penguin	razor	lamp	the adult penguin
2-ref13	square pan	round pan	pear	fridge	the square pan
2-ref14	fresh apple	rotten apple	comb	sausage	the fresh apple
2-ref15	dry stone	wet stone	cup	door	the dry stone
2-ref16	tall jug	short jug	kettle	panda	the tall jug
2-ref17	big cookie	small cookie	anchor	hammer	the big cookie

2-ref18	big cube	small cube	drum	vase	the big cube
2-ref19	small hat	big hat	shoe	cube	the small hat
2-ref20	small anchor	big anchor	drum	boat	the small anchor

Appendix B. Sample items for experiment 2a and 2b (adult version)

Item	Target	Competitor (contrast conditions only)			Utterance: *Pass me the X*
Under-informative					
und01	long skirt	short skirt	football	toilet	the skirt
und02	closed bag	open bag	strawberry	zebra	the bag
und03	old newspaper	recent newspaper	cherry	dog	the newspaper
und04	short sock	long sock	lobster	spoon	the sock
und05	stripy cup	spotty cup	tv	cucumber	the cup
und06	tall vase	short vase	key	bus	the vase
und07	new phone	old phone	hat	flower	the phone
und08	small star	big star	house	chick	the star
und09	sleeping baby	feeding baby	butterfly	tap	the baby
und10	full glass	empty glass	iron	broom	the glass
Over-informative					
over01	thin nail	fork	apple	cat	the thin nail
over02	tall boot	carrot	tree	microphone	the tall boot
over03	new phone	bike	compass	feather	the modern phone
over04	old computer	watch	guitar	mirror	the old computer
over05	intact egg	onion	camera	hairdryer	the unbroken egg
over06	closed umbrella	onion	ice-cream	stool	the closed umbrella
over07	open book	hamburger	fish	teddy	the open book
over08	unsliced bread	toothbrush	toad	pineapple	the unsliced bread
over09	unlit cigarette	tiger	plane	soap	the unlit cigarette
over10	modern rug	banana	crab	toaster	the modern rug
C: optimal + no contrast (opt1)					
opnocon01	banana	cube	fork	watch	the banana
opnocon02	comb	onion	hamburger	toothbrush	the comb
opnocon03	hammer	apple	tiger	tree	the hammer
opnocon04	pear	camera	ice-cream	guitar	the pear
opnocon05	sausage	toad	plane	crab	the sausage
opnocon06	boat	cat	microphone	feather	the boat
opnocon07	drum	stool	hairdryer	teddy	the drum
opnocon08	corkscrew	toaster	duck	soap	the corkscrew

opnocon09	flower	razor	anchor	drum	the flower
opnocon10	onion	cup	fridge	vase	the onion

D: optimal + contrast (opt2)

opcon01	tall jug	short jug	kettle	panda	the tall jug
opcon02	glass mug	china mug	duck	corkscrew	the glass mug
opcon03	adult penguin	baby penguin	razor	lamp	the adult penguin
opcon04	square pan	round pan	pear	fridge	the square pan
opcon05	fresh apple	rotten apple	comb	sausage	the fresh apple
opcon06	big cookie	small cookie	anchor	hammer	the big cookie
opcon07	big cube	small cube	drum	vase	the big cube
opcon08	big hat	small hat	shoe	cube	the small hat
opcon09	small anchor	big anchor	drum	boat	the small anchor
opcon10	dry stone	wet stone	cup	door	the dry stone

Fillers

Clefts

f1	big circle	small circle	cube	hammer	the big circle, pass me
f2	open door	closed door	carrot	cat	the open door, pass me
f3	cookie	tree	feather	compass	the cookie, pass me
f4	hairdryer	bike	microphone	mirror	the hairdryer, pass me
f5	Kettle	watch	guitar	onion	the kettle, pass me

Zero article

f6	short sock	long sock	camera	elephant	pass me short sock
f7	big bottle	small bottle	fish	hamburger	pass me big bottle
f8	feather	ice-cream	pineapple	toothbrush	pass me feather
f9	pineapple	banana	toilet	toad	pass me pineapple
f10	onion	tiger	plane	soap	pass me onion

Adjective–noun reversal

f11	plastic cup	paper cup	stool	teddy	pass me the cup plastic
f12	tall jug	small jug	toaster	strawberry	pass me the jug tall
f13	broken chair	intact chair	cherry	crab	pass me the chair broken
f14	closed eye	open eye	giraffe	spoon	pass me the eye closed
f15	wooden table	glass table	pencil	candle	pass me the table wooden

Scrambled word order

f16	small shoe	big shoe	chick	bus	me shoe pass small the
f17	big car	small car	house	key	me car pass big the
f18	teddy	cucumber	hotdog	iron	me teddy pass the
f19	soap	hat	tap	flower	me soap pass the
f20	watch	cake	football	dog	me watch pass the

References

Arts, A., 2004. Overspecification in instructive texts. Ph.D. Thesis, Tilburg University. Wolf Publishers, Nijmegen, The Netherlands. Bard, E.G., Robertson, D., Sorace, A., 1996. Magnitude estimation of linguistic acceptability. Language 72 (1), 32–68.

Brennan, S.E., Clark, H.H., 1996. Conceptual pacts and lexical choice in conversation. Journal of Experimental Psychology-Learning, Memory and Cognition 22 (6), 1482–1493.

Brown-Schmidt, S., Tanenhaus, M., 2006. Watching the eyes when talking about size: an investigation of message formulation and utterance planning. Journal of Memory and Language 54, 592–609.

Clark, H.H., Wilkes-Gibbs, D., 1986. Referring as a collaborative process. Cognition 22, 1–39.

Davies, C., Katsos, N., 2009. Are interlocutors as sensitive to over-informativeness as they are to under-informativeness? In: Poster presented at Pre-CogSci 2009, Workshop on the Production of Referring Expressions, Amsterdam.

Deutsch, W., Pechmann, T., 1982. Social-interaction and the development of definite descriptions. Cognition 11 (2), 159–184.

Dickson, W., 1982. Two decades of referential communication research: a review and meta-analysis. In: Brainerd, C.J., Presley, M. (Eds.), Verbal Processes in Children. Springer Verlag, New York, pp. 1–33.

Engelhardt, P.E., Bailey, K.G.D., Ferreira, F., 2006. Do speakers and listeners observe the Gricean Maxim of Quantity? Journal of Memory and Language 54 (4), 554–573.

Feeney, A., Scrafton, S., Duckworth, A., Handley, S.J., 2004. The story of some: everyday pragmatic inference by children and adults. Canadian Journal of Experimental Psychology 58 (2), 121–132.

Ford, W., Olson, D., 1975. The elaboration of the noun phrase in children's description of objects. Journal of Experimental Child Psychology 19 (3), 371–382.

Grice, P., 1975. Logic and conversation. In: Cole, P., Morgan, J.L. (Eds.), Syntax and Semantics. III. Speech Acts. Academic Press, New York, NY, pp. 41–58. Reprinted in Grice, P., 1989. Studies in the Way of Words. Harvard University Press, Cambridge, MA.

Guasti, M.T., Chierchia, G., Crain, S., Foppolo, F., Gualmini, A., Meroni, L., 2005. Why children and adults sometimes, but not always. compute implicatures. Language and Cognitive Processes 20 (5), 667–696.

Hendriks & Koster, this issue.

Hirschberg, J., 1991. A Theory of Scalar Implicature. Garland, New York.

Horn, L., 1984. Toward a new taxonomy for pragmatic inference. In: Schiffrin, D. (Ed.), Meaning, Form and Use in Context: Linguistic Applications, Proceedings of GURT'84. Georgetown University Press, Washington, DC, pp. 11–42.

Huang, Y., Snedeker, J., 2009. Semantic meaning and pragmatic interpretation in 5-year-olds: evidence from real-time spoken language comprehension. Developmental Psychology 45 (6), 723–1739.

Hurewitz, F., Papafragou, A., Gleitman, L., Gelman, R., 2006. Asymmetries in the acquisition of numbers and quantifiers. Language Learning and Development 2, 77–96.

Karmiloff-Smith, A., 1979. A Functional Approach to Child Language. C.U.P., Cambridge.

Katsos, N., 2009. Evaluating under-informative utterances with context-dependent and context-independent scales: experimental and theoretical implications. In: Sauerland, U., Yatsushiro, K. (Eds.), Experimental Semantics and Pragmatics, Palgrave Studies in Pragmatics, Language & Cognition.

Katsos, N., Bishop, D.V.M., submitted for publication. The development of informativeness from a speaker's and a comprehender's perspective. Paper presented at the XI Congress of the International Association for the Study of Child Language, 28 July–August 2008, Edinburgh, UK. An extended version is submitted to Cognition.

Katsos, N., Smith, N., 2010. Pragmatic Tolerance or a speaker–comprehender asymmetry in the acquisition of informativeness? In: Franich, K., Iserman, K.M., Keil, L.L. (Eds.), Proceedings of the 34th Annual Boston Conference in Language Development, Cascadilla Press, MA, USA.

Korkman, M., Kirk, U., Kemp, S., 1999. NEPSY: A Developmental Neuropsychological Assessment. The Psychological Corporation, San Antonio, TX. Maes, A., Arts, A., Noordman, L., 2004. Reference management in instructive discourse. Discourse Processes 37 (2), 117–144.

Mangold, R., Pobel, R., 1988. Informativeness and instrumentality in referential communication. Journal of Language and Social Psychology 7, 181–191. Matthews, D., Lieven, E., Tomasello, M., 2007. How toddlers and preschoolers learn to uniquely identify referents for others: a training study. Child Development 78, 1744–1759.

Noveck, I.A., 2001. When children are more logical than adults. Cognition 86, 253–282.

Papafragou, A., Musolino, J., 2003. Scalar implicatures: experiments at the semantics/pragmatics interface. Cognition 86, 253–282. Papafragou, A., Tantalou, N., 2004. Children's computation of implicatures. Language Acquisition 12 (1), 71–82.

Pechmann, T., 1989. Incremental speech production and referential overspecification. Linguistics 27 (1), 89–110.

Pouscoulous, N., Noveck, I., Politzer, G., Bastide, A., 2007. Processing costs and implicature development. Language Acquisition 14, 347–376. Reinhart, T., 2004. The processing cost of reference-set computation: acquisition of stress shift and focus. Language Acquisition 12, 109–155. Reinhart, T., 2006. Interface Strategies: Optimal and Costly Computations. MIT Press, Boston, MA.

Sedivy, J.C., 2002. Invoking discourse-based contrast sets and resolving syntactic ambiguities. Journal of Memory and Language 46 (2), 341–370. Sonnenschein, S., 1982. The effects of redundant communication on listeners—when more is less. Child Development 53 (3), 717–729. Whitehurst, G.J., 1976. Development of communication—changes with age and modeling. Child Development 47 (2), 473–482.