The role of salience in young children's processing of ad-hoc implicatures

Erica J. Yoon[1] & Michael C. Frank[1]

[1] Stanford University

Author Note

We would like to acknowledge Asher Kaye, Stephanie Hsiang, and Jacqueline Quirke for their assistance in data collection, and thank the staff and families at Children's Discovery Museum of San Jose and Bing Nursery School. All data, analysis code, and experiment files and links are available at https://github.com/ejyoon/simpimp_rs. This work was supported by a Postgraduate Doctoral Fellowship provided to EJY by Natural Sciences and Engineering Research Council of Canada, NSF #1456077, and Jacobs Advanced Research Fellowship to MCF.

Correspondence concerning this article should be addressed to Erica J. Yoon, Department of Psychology, Jordan Hall, 450 Serra Mall (Bldg. 420), Stanford, CA, 94305. E-mail: ejyoon@stanford.edu

Abstract

Language comprehension often requires making *implicatures*. For example, inferring that "I ate some of the cookies" implicates the speaker ate some *but not all* (scalar implicatures); and "I ate the chocolate-chip cookies" where there are both chocolate chip cookies and raisin cookies in the context implicates that the speaker ate the chocolate chip, but *not both the chocolate chip and raisin cookies* (ad-hoc implicatures). Children's ability to make scalar implicatures develops around age five, with ad-hoc implicatures emerging somewhat earlier. In the current work, using a time-sensitive tablet paradigm, we examined developmental gains in children's ad-hoc implicature processing, and found evidence for successful implicature computation by children as young as 3 years in a supportive context and substantial developmental gains in implicature computation from 2 to 5 years. We also tested whether one cause of younger children (2-year-olds)'s consistent failure to make implicatures is their difficulty in inhibiting an alternative interpretation that is more salient than the target meaning (the *salience hypothesis*). Our findings support this hypothesis: Younger children's failures with implicatures are likely related to effects of the salience mismatch between possible interpretations.

Keywords: Pragmatics; cognitive development; language processing; implicature; tablet

Word count: 6717

The role of salience in young children's processing of ad-hoc implicatures

Introduction

Language comprehension often requires inferring an intended meaning that goes beyond the literal semantics of what a speaker says. In Grice (1975)'s account, conversation is a cooperative act: Speakers choose utterances such that the listener can understand the intended message, and listeners in turn interpret these utterances with the assumption of the speaker's cooperativeness in mind. For example, expecting a cooperative speaker to have produced a maximally informative utterance for the present conversational needs, the listener can make inferences that go beyond the literal meanings of the speaker's words.

The non-literal interpretations computed through these inferential processes are called pragmatic implicatures. For example, "I ate some of the cookies" implicates that the speaker ate some *but not all of the cookies*, because a cooperative speaker who ate all of them would have said "all," which is more informative than the alternative "some." This inference is an example of a *scalar implicature*, in which the use of a weaker proposition ("some ~") leads the interpreter to believe the negation of a proposition with a stronger meaning ("all ~"). Another kind of implicature, *ad-hoc implicature*, is context-based: "I ate the chocolate chip cookies" in a context where two kinds of cookies – chocolate chip and raisin – are available, implicates that the speaker ate the chocolate chip *but not both the chocolate chip cookies and raisin cookies*. In this case, the context sets up a contrast between the proposition offered ("ate the chocolate chip cookies") and the stronger alternative to be negated ("ate the chocolate chip and raisin cookies")[1].

Implicatures like these have been an important case study for pragmatics more broadly.

[1] Grice (1975) calls these implicatures generalized (scalar) vs. particularized (ad-hoc), but we use a theory-neutral designation here. An alternative analysis of the second implicature relies on the contrast between "the chocolate chip cookies" and "the cookies" – since the second entails the first, there is an implicature. For our purposes, this entailment relation is still ad-hoc in the sense that "the cookies" does not entail "the chocolate chip cookies" in discourse contexts in which no chocolate chip cookies are part of common ground.

Both classic theories of communication (e.g., Sperber & Wilson, 1995) and more recent probabilistic models of pragmatic inference (e.g., Frank & Goodman, 2012; see Goodman & Frank, 2016 for review) describe the processes that language users use to compute such implicatures. And a rich psycholinguistic literature has measured adults' processing of implicatures relative to literal interpretations and found that adults robustly compute implicatures, albeit sometimes more slowly than unambiguous literal meanings (Bott, Bailey, & Grodner, 2012; Breheny, Ferguson, & Katsos, 2013; Huang & Snedeker, 2009a). How does the ability to make implicatures develop? Since implicature computation is an important indicator of broader pragmatic understanding, many studies have tested children's abilities. In these experiments, children tend to have the most difficulty with scalar implicatures relying on quantifiers, modals, and other functional elements. For example, in Papafragou and Musolino (2003)'s study, a puppet saw three out of three horses jump over a fence, and described the scene infelicitously by saying "Some of the horses jumped over the fence." Adults tend to reject this infelicitous statement, whereas 5-year-old children mostly accept it, suggesting that children failed to compute the relevant scalar implicature (though see Katsos & Bishop, 2011, for an alternative explanation). Besides struggling with *some* vs. *all* (Huang & Snedeker, 2009b; Hurewitz, Papafragou, Gleitman, & Gelman, 2006; Noveck, 2001), children in the same age range have consistently failed to compute implicatures involving scalar contrasts, including *a* vs. *some* (Barner, Chow, & Yang, 2009), *might* vs. *must* (Noveck, 2001), and *or* vs. *and* (Chierchia, Crain, Guasti, Gualmini, & Meroni, 2001).

While children struggle on many scalar implicature tasks, they tend to be more successful at computing ad-hoc implicatures (which depend on context, rather than lexical scales). One potential difficulty in a typical scalar implicature task is the need to generate relevant alternatives to a given scalar term. For children to hear "some of the horses jumped over the fence" and derive the implicature "some *but not all*," they must first realize that "all" is the relevant alternative to "some." Barner, Brooks, and Bale (2011) argued that children's failures in scalar implicature tasks are due to their lack of ability to generate the alternative to negate spontaneously upon hearing the term offered. Barner et al. (2011)'s claim predicts that children's implicature computation should improve when they can access the relevant alternatives. Consistent with this claim, children can be primed with relevant scalar alternatives, leading to enhanced implicature performance (Skordos & Papafragou, 2016). Furthermore, children show substantially improved implicature computation in ad-hoc implicature tasks – which provided access to relevant alternatives in context – compared to scalar implicature tasks (Horowitz, Schneider, & Frank, in press; Katsos & Bishop, 2011; Papafragou & Tantalou, 2004; Stiller, Goodman, & Frank, 2015).

For example, Stiller et al. (2015) showed 2.5- to 5-year-old children three different faces: a face with no item; a face with only glasses; and a face with glasses and a top-hat, and asked children to choose one of the three faces as the referent in a puppet's statement, "My friend has glasses." In this task, there was no need for children to spontaneously generate the alternative ("glasses and hat") to the term used ("glasses") because the alternative was visible in the context. Children as young as 3.5 years chose the face with only glasses as the referent, suggesting that they successfully computed the implicature that the puppet's friend has "glasses but not both glasses and hat." Similarly, in one study that tested both scalar and ad-hoc implicature computation, 4-year-olds successfully made ad-hoc implicatures, but performed poorly on scalar implicatures using the same stimuli (Horowitz et al., in press).

Despite older children's success, children below 3 years of age appear to struggle with even simple ad-hoc implicatures. Even in the ad-hoc paradigm described above (Stiller et al., 2015), 2.5- and 3-year-olds still did not make the implicature-consistent choice at above-chance levels. Does this finding imply that young toddlers lack pragmatic understanding, specifically an awareness of the need for informativeness in cooperative communication? On the contrary, children are sensitive to informativeness in communication: From age two onward, when they are asked to produce referring expressions, children appear to recognize the level of referential ambiguity of their own expressions and attempt to provide more information through speech and gestures in more ambiguous situations (e.g., instead of "the boy," saying "the boy with the dog"; or naming an object while pointing in cases where the point alone is not precise enough; Matthews, Butcher, Lieven, & Tomasello, 2012; O'Neill & Topolovec, 2001). Hence, a lack of sensitivity to the need for communicative informativeness does not seem to be the problem for toddlers' implicature processing. So what causes toddlers' failures in implicature tasks?

One potential explanation for younger children's struggle with ad-hoc implicatures is the mismatch in salience between potential interpretations. For example, in Stiller et al. (2015)'s study, a target referent (e.g., face with glasses only) had fewer features than its alternative distractor to be rejected (e.g., face with glasses and hat). The distractor, which had a greater number of nameable features, was more salient both perceptually and conceptually, likely drawing children's attention more strongly than the target. This kind of task may be challenging to children because their executive function is not yet fully developed (Davidson, Amso, Anderson, & Diamond, 2006; Diamond & Taylor, 1996) – specifically their ability to inhibit responses to salient targets (but see Discussion for further consideration of whether children's failures should be attributed to their inhibitory control abilities per se). Further, issues in referent selection tasks may reflect analogous problems in naturalistic language comprehension for children, in which the goal is often to figure out what referent a speaker is talking about. Thus, the salience account might apply to pragmatic inferences in real-world language comprehension as well.

This asymmetry between correct but weaker target meaning and incorrect but more salient distractor meaning is present in other types of implicatures too, though less obviously so. Scalar implicature is typically described as rejecting the term that yields the stronger propositional meaning (e.g., "all" of the cookies) and adding its negation to the weaker proposition (e.g., "some but not all" of the cookies). Computing a forced-choice scalar implicature thus also requires avoiding the stronger meaning, which typically describes a larger set size (all of the cookies). Although the referents in such tasks are not always pictured visually side-by-side, they are in at least some paradigms (e.g., Huang & Snedeker, 2009b). Such issues could further exacerbate the difficulties of scalar implicature, at least for some age groups. We return in the Discussion to the question of whether distractor salience could plausibly explain some of the data on scalar implicature development.[2]

For our experiment, we adopted a referent selection method, in which participants were asked to select a referent among a set of candidates. As mentioned earlier, referent selection paradigms have shown evidence of successful implicature computation in youngest children to date (Horowitz et al., in press; Stiller et al., 2015), and are analogous to the task of language comprehension in naturalistic language environments: identifying a speaker's intended referent. We implemented the referent selection method using a tablet paradigm to examine children's reaction times for selecting the target referent (Frank, Sugarman, Horowitz, Lewis, & Yurovsky, 2016). Compared to previous studies, we reduced the number of potential referents in context to further simplify the task: In Stiller et al. (2015)'s paradigm, there were three potential referents in the context (face with no item, face with only glasses, face with glasses and hat); in our current paradigm, we presented two instead of three potential referents (e.g. plate with a carrot and plate with a carrot and a banana) to minimize cognitive load for children.

We present data from two independent samples: The first planned sample of children across four age groups (2-, 3-, 4-, and 5-year-olds) initially showed a pattern consistent with the salience hypothesis, where children were more accurate for trials with lower salience contrasts than for trials with higher salience contrasts. This effect was relatively small, however, and our analysis plan was not prespecified, leading us to worry about the possibility that analytic flexibility might have led us to overestimate our effect (e.g., Simmons, Nelson, & Simonsohn, 2011). We thus collected a second, fully preregistered sample of children

[2] As we discuss below, this account is not mutually exclusive with the alternatives hypothesis described above. Indeed, both are likely true and likely contribute to children's difficulty with implicatures to different degrees in different tasks and at different ages. For more complex alternative sets, the challenge may primarily be identifying the appropriate alternatives, while for simpler alternatives, difficulties may lie primarily in overcoming the pull of the stronger one.

across the three youngest groups (2-, 3- and 4-year-olds) to replicate this initial finding.

Methods

We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study.

Participants

In the original sample, either parents and their children visiting Children's Discovery Museum (San Jose, CA), or children in a local nursery school were invited to participate in a tablet study, and a total of 123 children were recruited. Participants were excluded from the sample for the following reasons: age other than 2 to 5 years (n = 3); parent-reported English exposure less than our prespecified criterion of 75% (n = 5); parental interference (n = 2); and noncompliance or difficulty with the experimental procedure (n = 9). After excluding participants who completed fewer than the prespecified number of 10 trials (n = 2), the final sample consisted of 102 children (see Table 1).

In the replication sample, a total of 116 children were recruited, all at Children's Discovery Museum in San Jose. Reasons for exclusions were: age other than 2 to 4 years (n = 11); parent-reported English exposure less than our prespecified criterion of 75% (n = 15); parental interference (n = 3); noncompliance or difficulty with the experimental procedure (n = 3); and technical error (n = 4). The final sample consisted of 80 children (no participant was excluded for completing fewer than 10 trials).

Stimuli and Design

On each trial, participants saw two images: a target and distractor, which could either be an item with a single feature (e.g. a lunchbox with only an apple or only an orange), or an item with double features (e.g., a lunchbox with an apple and an orange). In each trial, a pre-recorded voice said a sentence (e.g., "Look at these lunchboxes. Elmo's lunchbox has an apple."). After participants chose the object that was being referred to, a green box appeared around the chosen object to show that the choice had been made. For each trial, we recorded the participant's accuracy, or whether he or she selected the correct target referent, and reaction time, or time spent between naming of the referent ("…an *apple*") and the participant's referent selection.

There were three types of test trials (shown at the top of each panel in Figure 1). In *implicature* trials, the target item had a single feature (e.g., an apple), and the distractor item had two or three features (see below for the manipulation of number of features) – one that was in common with the target (e.g., an apple) and the other feature(s) that was/were unique (e.g., an orange). The test sentence named the feature that was common to the target and distractor. Thus, if participants understood that "Elmo's lunchbox has an apple" implicates "Elmo's lunchbox has an apple *but not an orange*" in the given context, it was predicted that they would choose the target more often than the distractor; otherwise, if they did not make implicatures, they would choose the two at equal rates (or even choose the distractor more often depending on the degree of saliency contrast – see below).

There were two additional trial types, with semantically unambiguous targets: *Control-double* trials looked identical to implicature trials, but the target and distractor were switched, such that the double-feature item was the target and the single-feature item was the distractor, and the test sentence named the unique feature on the target. *Control-single* trials presented two items that each had a unique single feature, and either could be the target. Children saw 4 implicature, 4 control-double, and 8 control-single trials; adults saw 6 implicature, 6 control-double, and 12 control-single trials.

Each trial type was further divided by the number of features present on the target and distractor (shown on the right side of Figure 1): Within implicature trials, *fewer-feature* (2-vs-1) trials presented two features (an apple and an orange) on the distractor and one feature (an apple) on the target, whereas *more-feature* (3-vs-1) trials presented three features (an apple, an orange, and a cookie) on the distractor and one feature on the target; Within control-double trials, *fewer-feature* (2-vs-1) trials presented two features (an apple and an orange) on the target and one feature (an apple) on the distractor, whereas *more-feature* (3-vs-1) trials presented three features (an apple, an orange, and a cookie) on the distractor and one feature on the target; Lastly, within control-single trials, *fewer-feature* (1-vs-1) trials presented one feature each on the distractor and the target, whereas *more-feature* (2-vs-2) trials presented two features each on the distractor and on the target. We hypothesized that older children would choose the target more often in the more-feature implicature trials than the fewer-feature implicature trials due to the strengthening of implicatures, and that, on the contrary, younger children would choose the target less often in the more-feature trials than the fewer-feature trials due to increased saliency of the distractor.

There were six sets of item and feature types, and the features were named with nouns found on the MacArthur-Bates Communicative Development Inventory word list (Fenson et al., 1994). Two orders of the test trials were created, such that trial types and item types were counterbalanced and trial order was pseudo-randomized across the two orders.

Procedure

An experimenter introduced children to the task as a game on a tablet. Then they completed two practice trials, where they were asked to select an obvious, unambiguous referent (e.g., "cow" as opposed to "rabbit"), followed by 16 test trials.

Data analysis

We used R (Version 3.4.3; R Core Team, 2017) and the R-packages *bindrcpp* (Version 0.2; Müller, 2017a), *brms* (Version 2.0.1; Bürkner, 2017), *dplyr* (Version 0.7.4; Wickham, Francois, Henry, & Müller, 2017), *forcats* (Version 0.2.0; Wickham, 2017a), *ggplot2* (Version 2.2.1; Wickham, 2009), *ggthemes* (Version 3.4.0; Arnold, 2017), *here* (Version 0.1; Müller, 2017b), *langcog* (Version 0.1.9001; Braginsky, Yurovsky, & Frank, n.d.), *lme4* (Version 1.1.15; D. Bates, Mächler, Bolker, & Walker, 2015), *Matrix* (Version 1.2.12; D. Bates & Maechler, 2017), *papaja* (Version 0.1.0.9655; Aust & Barth, 2017), *purrr* (Version 0.2.4; Henry & Wickham, 2017), *Rcpp* (Eddelbuettel & Balamuta, 2017; Version 0.12.14; Eddelbuettel &

François, 2011), *readr* (Version 1.1.1; Wickham, Hester, & Francois, 2017), *stringr* (Version 1.2.0; Wickham, 2017b), *tibble* (Version 1.3.4; Müller & Wickham, 2017), *tidyr* (Version 0.7.2; Wickham & Henry, 2017), *tidyverse* (Version 1.2.1; Wickham, 2017c), and *xtable* (Version 1.8.2; Dahl, 2016) for all our analyses.

Results

We were interested in children's processing of implicatures in comparison to unambiguous utterances, and developmental gains across ages. We used two different measures: (1) accuracy and (2) reaction time for choosing the correct referent. For each measure, we asked: (a) do children show developmental gains in selection of the target referent? And (b) does children's performance vary depending on salience contrast? That is, when there are a relatively greater number of features on the distractor, do children have more difficulty and are they slower in choosing the correct referent?

As per our standard operating procedures, we removed trials in which the log of reaction time was more than 3 standard deviations above or below the mean (upper bound: 14.04 seconds; lower bound: 0.47 second; percentage of data excluded: 1.67 %). Throughout this section, we used Bayesian linear mixed-effects models (`brms` package in R; Bürkner, 2017) using crossed random effects of participant, item, and sample (original vs. replicationn) with the maximal random effects structure supported by the design (Barr, Levy, Scheepers, & Tily, 2013; A. Gelman & Hill, 2006). Age is plotted in year bins, but was analyzed as a continuous variable, scaled and centered, in our statistical model.

Accuracy

The analysis of the accuracy rate (Figure 2) showed that children across all ages were able to identify the target in control trials, indicating that, as expected, they can readily compute unambiguous meanings. In implicature trials, 4- and 5-year-olds' performances were nearly at ceiling, replicating the previous results (Horowitz et al., in press; Stiller et al., 2015). In our paradigm, even 3-year-olds chose the inferential target above chance[3] (original sample: $t(58) = 6.82$, $p < 0.001$; replication sample: $t(57) = 5.33$, $p < 0.001$). On the other hand, 2-year-olds' performance in implicature trials did not differ from chance overall, but their performance varied depending on the number of features present. In 3-vs-1 trials (i.e., with a relatively greater number of features on the distractor), 2-year-olds did not choose the correct target referent, and even tended to choose the distractor somewhat more often numerically (original sample: $t(23) = -1.42$, $p = 0.17$; replication sample: $t(24) = -0.72$, $p = 0.48$). However, In 2-vs-1 trials (with fewer features on the distractor), 2-year-olds tended to choose the target more often than the distractor. This difference was numerically present in both samples and statistically significant in one (original sample: $t(26) = 0.46$, $p = 0.65$; replication sample: $t(24) = 2.57$, $p = 0.02$). By 4 years, this difference in accuracy rate between 2-vs-1 and 3-vs-1 trials was not present.

A Bayesian linear mixed-effects model predicting accuracy based on age, trial type and number of features (salience contrast; more-feature vs. fewer-feature) showed a three-way positive interaction of age, implicature trials, and number of features (Table 2). Thus, unlike control trials, in which children's performances did not differ by salience contrast, implicature trials showed lower accuracy in 3-vs-1 than 2-vs-1 trials in younger children, but not in older children. This result supports our initial hypothesis that salience contrast may lead to a greater struggle for younger children with the implicature task due to a higher demand for inhibiting response to distractor with greater salience.

[3]Because our task is a two-alternative forced choice, we define chance to be 50% across all trials. This baseline is a standard comparison that reflects the possibility that a child was completely inattentive to the task and chose completely at random. This baseline is more conservative than a salience-based baseline, which would likely suggest that the correct (inferentially-consistent) target would be chosen less than 50% of the time (e.g., the "mumble" condition in Stiller et al., 2015).

Reaction time

With increasing age, children computed both implicatures and unambiguous meanings and identified the target faster (Figure 3). A Bayesian linear mixed-effects model predicting reaction time based on age, trial type and number of features present showed a positive two-way interaction between age and implicature trial (Table 3), indicating that the speed of implicature computation did not improve with age as much as the speed of processing unambiguous meanings. Together with the accuracy finding, this result suggests that though children become proficient at determining the *correct* target referents for ad-hoc implicatures by 5 years, implicature processing develops relatively more slowly.

We also observed a positive two-way interaction between control-double trials and number of features, indicating that children took longer to identify the target in control-double trials with more features than in control-single trials with more features. Interestingly, there was no interaction between inference trials and number of features, or between inference trial, age and number of features. Why would this be? We did not have a pre-specified hypothesis regarding this pattern of data, but we speculate that once a feature is named (e.g., Elmo's lunchbox has an apple), it is relatively easier to find the feature in an inferential target image than in the distractor image. The target feature is by itself in the target referent, whereas it is grouped with with other features in the distractor. Thus, the inference trials may allow easy perceptual access to the target feature but also competition with the overall perceptual salience of the distractor. These factors might cancel one another out and lead to undifferentiated reaction times and hence the lack of reaction time interactions.

Discussion

In our experiment, we confirmed 3- to 5-year-old children's successes on ad-hoc implicature computation, and saw substantial developmental gains in their accuracy and speed. 4- and 5-year-old children successfully computed ad-hoc implicatures and identified the inferential targets, consistent with previous findings. We found evidence of successful implicature computation even in 3-year-olds. Between 2 and 5 years, there was a clear improvement in processing skills with increasing age, such that correct referent identification was more accurate and faster across both control and implicature trials. Thus, these findings add to the existing literature to attest to children's growing proficiency in pragmatic processing.

We also investigated the salience hypothesis, namely that one cause of young children's struggle with implicatures stems from their difficulty to inhibit choosing the more salient distractor. In earlier work, there was some numerical suggestion of 2-year-olds' preference for the more salient but pragmatically incorrect distractor (Stiller et al., 2015). We predicted that increasing the salience of this distractor would result in decreased performance for younger children while increasing performance for older children. The first part of this prediction was clearly supported in our data, with younger children performing worse when the distractor was more salient. Although we observed numerical hints of a gain in accuracy for older children in one sample, we did not see a consistent facilitation effect. We suspect this finding is due to a ceiling effect: Referent selection via ad-hoc implicature is relatively trivial for four-year-olds (Horowitz et al., in press). However, we saw a possible age-related advantage of pragmatic strengthening in the speed of computation: Whereas younger children tended to be slower in trials with a greater number of features for both unambiguous and inferential meanings, older children began to close the gap and become faster to compute implicatures given increased distractor saliency.

Although our salience account is most manifest in the kind of simple referent selection tasks we used here, we believe it applies more broadly to implicature computation beyond the scope of these tasks. Any pragmatic implicature requires an asymmetry in the "strength" of the alternatives. In ad-hoc referent-selection contexts, the stronger (more salient) alternative is the item with more features. In scalar implicatures, the implicature that you ate *some but not all of the cookies* is only possible because there is a stronger alternative ("all"). It remains an open empirical question whether the salience mismatch account might explain children's difficulty with these other cases of implicatures as well.

One further application of our account is to word learning contexts, where children's learning of a novel word is facilitated when the target referent is more (not less) salient than its alternative. For example, Frank and Goodman (2014) used an analogous pragmatic inference paradigm in a word learning context: Participants heard a novel label (e.g., "a dinosaur with a *dax*") used to describe an object with two features (a dinosaur with a hat and a bandanna) in the presence of another dinosaur that had one but not the other of the features (a dinosaur with a hat only). 3- and 4-year-olds performed quite well in mapping the novel label to the unique (more informative) feature. In this paradigm, the novel label was being mapped to the more, rather than less, salient object. Similarly, in classic "mutual exclusivity" paradigms (Markman & Wachtel, 1988), by around 18 months, participants succeed in mapping a novel label to a novel object (Halberda, 2003). While the mechanisms underlying this empirical phenomenon are complex, it is well-established that the salience of the novel target is an important factor in children's success (see Markman, Wasow, & Hansen, 2003 for discussion). Overall, evidence for children's pragmatic word learning emerges earlier than implicature computation: Children succeed in these tasks substantially at earlier ages than even in our simplified implicature paradigm. Our account suggests that one reason for this asymmetry is because implicature tasks require selecting the *less* salient alternative while word learning tasks typically ask participants to select a *more* salient alternative.

Our findings help in the construction of a comprehensive developmental account of processing of implicatures, and pragmatic inferences in general. In the samples that have been studied in this literature, by 2 years of age, children begin to be aware that informativeness is important to communication. Nevertheless, our findings suggest that the salience contrasts inherent in many pragmatic situations may keep them from successfully processing implicatures. Further, these same factors are plausibly in play during pragmatic word learning tasks (Yurovsky & Frank, 2017). By 3 to 4 years, the ability to inhibit these salient targets is more developed, and they start to compute ad-hoc implicatures when relevant alternatives to the speaker's words are provided in context. Scalar implicature performance develops more slowly, however, as children's ability to access the relevant alternatives (and their semantics) is only beginning to emerge (Barner et al., 2011; Horowitz et al., in press; Skordos & Papafragou, 2016); their performance during these ages is highly variable and dependent on the nature of the context and its pragmatic demands (Papafragou & Tantalou, 2004).

One important challenge for this viewpoint is the nature of the ability that children use to overcome the pull of the salient alternative. One possible naive mapping for the ability would be to the broader construct of executive function, which undergoes substantial developmental changes during this period (Davidson et al., 2006; Diamond & Taylor, 1996). But executive function is a multi-faceted construct (Miyake et al., 2000), and the particular components that would be expected to predict visual (and perhaps conceptual) disengagement with a particular referent is unclear. Our own studies attempting to probe individual difference correlations between executive function and implicature ability in development have not been successful (e.g., Horowitz et al., in press; Nordmeyer, Yoon, & Frank, 2016). Thus, a target for future work is to better characterize the particular cognitive changes that relate to the developmental effects we have observed here.

There are several further limitations of our work here. First, our salience manipulation involved manipulation of the number of features present on an item, which might have caused a potential confound between salience and processing time. For example, children's greater looking to the distractor might have been caused by a real desire to acquire more information, rather than the mere perceptual salience of the distractors. Second, as with nearly all work in the literature on implicature processing, we address the performance of only relatively high socioeconomic status children in a Western context. In our ongoing work we address the generalizability of our task to other developmental contexts (Fortier, Kellier, Fernández Flecha, & Frank, in prep).

In sum, our work shows evidence that from at least 3 years, children are able to compute ad-hoc implicatures, and that younger children's failures with implicatures are likely related to effects of the salience mismatch between possible referents. This pattern is consistent with a broader generalization, namely that tasks that have typically been used to look at children's implicature processing have a variety of extraneous processing demands, which may explain why it has been difficult to see children's underlying pragmatic abilities in such paradigms. Thus, our work demonstrates the importance of using a range of methods to measure children's pragmatic processing.

References

Arnold, J. B. (2017). *Ggthemes: Extra themes, scales and geoms for 'ggplot2'*. Retrieved from https://CRAN.R-project.org/package=ggthemes

Aust, F., & Barth, M. (2017). *papaja: Create APA manuscripts with R Markdown*. Retrieved from https://github.com/crsh/papaja

Barner, D., Brooks, N., & Bale, A. (2011). Accessing the unsaid: The role of scalar alternatives in children's pragmatic inference. *Cognition*, *118*(1), 84–93. doi:10.1016/j.cognition.2010.10.010

Barner, D., Chow, K., & Yang, S.-J. (2009). Finding one's meaning: A test of the relation between quantifiers and integers in language development. *Cognitive Psychology*, *58*(2), 195–219. doi:10.1016/j.cogpsych.2008.07.001

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

Bates, D., & Maechler, M. (2017). *Matrix: Sparse and dense matrix classes and methods*. Retrieved from https://CRAN.R-project.org/package=Matrix

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi:10.18637/jss.v067.i01

Bott, L., Bailey, T. M., & Grodner, D. (2012). Distinguishing speed from accuracy in scalar implicatures. *Journal of Memory and Language*, *66*(1), 123–142. doi:10.1016/j.jml.2011.09.005

Braginsky, M., Yurovsky, D., & Frank, M. C. (n.d.). *Langcog: Language and cognition lab things*. Retrieved from http://github.com/langcog/langcog

Breheny, R., Ferguson, H. J., & Katsos, N. (2013). Taking the epistemic step: Toward a model of on-line access to conversational implicatures. *Cognition*, *126*(3), 423–440. doi:10.1016/j.cognition.2012.11.012

Bürkner, P.-C. (2017). brms: An R package for bayesian multilevel models using Stan.

Journal of Statistical Software, *80*(1), 1–28. doi:10.18637/jss.v080.i01

Chierchia, G., Crain, S., Guasti, M. T., Gualmini, A., & Meroni, L. (2001). The acquisition of disjunction: Evidence for a grammatical view of scalar implicatures. In *Proceedings of BUCLD 25* (Vol. 157, p. 168).

Dahl, D. B. (2016). *Xtable: Export tables to latex or html*. Retrieved from https://CRAN.R-project.org/package=xtable

Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia*, *44*(11), 2037–2078. doi:10.1016/j.neuropsychologia.2006.02.006

Diamond, A., & Taylor, C. (1996). Development of an aspect of executive control: Development of the abilities to remember what I said and to "do as I say, not as I do". *Developmental Psychobiology*, *29*, 315–334. doi:10.1002/(SICI)1098-2302(199605)29:4\%3C315::AID-DEV2\%3E3.3.CO;2-C

Eddelbuettel, D., & Balamuta, J. J. (2017). Extending extitR with extitC++: A Brief Introduction to extitRcpp. *PeerJ Preprints*, *5*, e3188v1. doi:10.7287/peerj.preprints.3188v1

Eddelbuettel, D., & François, R. (2011). Rcpp: Seamless R and C++ integration. *Journal of Statistical Software*, *40*(8), 1–18. doi:10.18637/jss.v040.i08

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., … Stiles, J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, i–185. doi:10.2307/1166093

Fortier, M., Kellier, D., Fernández Flecha, M., & Frank, M. C. (in prep). Ad-hoc pragmatic implicatures among shipibo-konibo children in the peruvian amazon. doi:10.17605/OSF.IO/X7AD9

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games.

Science, *336*(6084), 998–998. doi:10.1016/j.cogpsych.2014.08.002

Frank, M. C., & Goodman, N. D. (2014). Inferring word meanings by assuming that speakers are informative. *Cognitive Psychology*, *75*, 80–96. doi:10.1016/j.cogpsych.2014.08.002

Frank, M. C., Sugarman, E., Horowitz, A. C., Lewis, M. L., & Yurovsky, D. (2016). Using tablets to collect data from young children. *Journal of Cognition and Development*, *17*, 1–17. doi:10.1080/15248372.2015.1061528

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press.

Goodman, N. D., & Frank, M. C. (2016). Pragmatic language interpretation as probabilistic inference. *Trends in Cognitive Sciences*, *20*(11), 818–829.

Grice, H. P. (1975). Logic and conversation. *Syntax and Semantics*, *3*, 41–58.

Halberda, J. (2003). The development of a word-learning strategy. *Cognition*, *87*(1), B23–B34.

Henry, L., & Wickham, H. (2017). *Purrr: Functional programming tools*. Retrieved from https://CRAN.R-project.org/package=purrr

Horowitz, A. C., Schneider, R. M., & Frank, M. C. (in press). The trouble with quantifiers: Explaining children's deficits in scalar implicature. *Child Development*.

Huang, Y. T., & Snedeker, J. (2009a). Online interpretation of scalar quantifiers: Insight into the semantics–pragmatics interface. *Cognitive Psychology*, *58*(3), 376–415. doi:10.1016/j.cogpsych.2008.09.001

Huang, Y. T., & Snedeker, J. (2009b). Semantic meaning and pragmatic interpretation in 5-year-olds: Evidence from real-time spoken language comprehension. *Developmental Psychology*, *45*(6), 1723. doi:10.1037/a0016704

Hurewitz, F., Papafragou, A., Gleitman, L., & Gelman, R. (2006). Asymmetries in the acquisition of numbers and quantifiers. *Language Learning and Development*, *2*(2), 77–96. doi:10.1207/s15473341lld0202_1

Katsos, N., & Bishop, D. V. (2011). Pragmatic tolerance: Implications for the acquisition of informativeness and implicature. *Cognition*, *120*(1), 67–81. doi:10.1016/j.cognition.2011.02.015

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, *20*(2), 121–157.

Markman, E. M., Wasow, J. L., & Hansen, M. B. (2003). Use of the mutual exclusivity assumption by young word learners. *Cognitive Psychology*, *47*(3), 241–275.

Matthews, D., Butcher, J., Lieven, E., & Tomasello, M. (2012). Two-and four-year-olds learn to adapt referring expressions to context: Effects of distracters and feedback on referential communication. *Topics in Cognitive Science*, *4*(2), 184–210. doi:10.1111/j.1756-8765.2012.01181.x

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100.

Müller, K. (2017a). *Bindrcpp: An 'rcpp' interface to active bindings*. Retrieved from https://CRAN.R-project.org/package=bindrcpp

Müller, K. (2017b). *Here: A simpler way to find your files*. Retrieved from https://CRAN.R-project.org/package=here

Müller, K., & Wickham, H. (2017). *Tibble: Simple data frames*. Retrieved from https://CRAN.R-project.org/package=tibble

Nordmeyer, A. E., Yoon, E. J., & Frank, M. C. (2016). Distinguishing processing difficulties in inhibition, implicature, and negation. In *Proceedings of the 38th annual meeting of the cognitive science society*.

Noveck, I. A. (2001). When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition*, *78*(2), 165–188. doi:10.1016/S0010-0277(00)00114-1

O'Neill, D. K., & Topolovec, J. C. (2001). Two-year-old children's sensitivity to the referential (in) efficacy of their own pointing gestures. *Journal of Child Language*, *28*(1), 1–28. doi:10.1017/S0305000900004566

Papafragou, A., & Musolino, J. (2003). Scalar implicatures: Experiments at the semantics–pragmatics interface. *Cognition*, *86*(3), 253–282. doi:10.1016/S0010-0277(02)00179-8

Papafragou, A., & Tantalou, N. (2004). Children's computation of implicatures. *Language Acquisition*, *12*(1), 71–82. doi:10.1207/s15327817la1201_3

R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366.

Skordos, D., & Papafragou, A. (2016). Children's derivation of scalar implicatures: Alternatives and relevance. *Cognition*, *153*, 6–18. doi:10.1016/j.cognition.2016.04.006

Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition*. Oxford: Blackwell.

Stiller, A., Goodman, N. D., & Frank, M. C. (2015). Ad-hoc implicature in preschool children. *Language Learning and Development*. doi:10.1080/15475441.2014.927328

Wickham, H. (2009). *Ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. Retrieved from http://ggplot2.org

Wickham, H. (2017a). *Forcats: Tools for working with categorical variables (factors)*. Retrieved from https://CRAN.R-project.org/package=forcats

Wickham, H. (2017b). *Stringr: Simple, consistent wrappers for common string operations*. Retrieved from https://CRAN.R-project.org/package=stringr

Wickham, H. (2017c). *Tidyverse: Easily install and load the 'tidyverse'*. Retrieved from https://CRAN.R-project.org/package=tidyverse

Wickham, H., & Henry, L. (2017). *Tidyr: Easily tidy data with 'spread()' and 'gather()' functions*. Retrieved from https://CRAN.R-project.org/package=tidyr

Wickham, H., Francois, R., Henry, L., & Müller, K. (2017). *Dplyr: A grammar of data manipulation*. Retrieved from https://CRAN.R-project.org/package=dplyr

Wickham, H., Hester, J., & Francois, R. (2017). *Readr: Read rectangular text data*. Retrieved from https://CRAN.R-project.org/package=readr

Yurovsky, D., & Frank, M. C. (2017). Beyond naive cue combination: Salience and social cues in early word learning. *Developmental Science*. doi:10.1111/desc.12349

Table 1

Demographic information of participants in the original and replication samples.

Sample	Age bin	Number of participants	Mean (years)	SD (years)	% Girls
original	2	27	2.51	0.31	70.40
original	3	30	3.54	0.28	56.70
original	4	26	4.45	0.29	34.60
original	5	19	5.30	0.23	57.90
replication	2	25	2.66	0.27	56.00
replication	3	29	3.49	0.27	55.20
replication	4	25	4.39	0.29	40.00

Table 2

Predictor mean estimates with standard deviation and 95% credible interval information for a Bayesian linear mixed-effects model predicting accurate selection of target.

Predictor	Mean	SD	95% CI-Lower	95% CI-Upper
Intercept	4.60	3.86	-4.94	11.71
Age	2.30	3.80	-5.84	10.06
Control-double	-0.07	3.53	-7.86	7.71
Implicature	-3.33	4.41	-13.11	4.46
More features	-0.68	3.70	-9.83	5.81
Control-double * Age	-0.72	0.38	-1.46	0.03
Implicature * Age	-1.00	0.38	-1.77	-0.29
More features * Age	-0.59	0.35	-1.28	0.09
Control-double * More features	-0.06	0.56	-1.16	1.02
Implicature * More features	0.23	0.63	-1.00	1.46
Control-double * Age * More features	0.66	0.41	-0.13	1.49
Implicature * Age * More features	1.16	0.44	0.32	2.02

Table 3

Predictor mean estimates with standard deviation and 95% credible interval information for a Bayesian linear mixed-effects model predicting log reaction time to select the target.

Predictor	Mean	SD	95% CI-Lower	95% CI-Upper
Intercept	7.80	2.24	3.19	12.79
Age	-0.22	2.83	-6.43	5.97
Control-double	0.02	3.16	-6.98	6.05
Implicature	-0.05	3.65	-8.44	6.01
More features	0.25	2.67	-4.92	6.56
Control-double * Age	-0.03	0.02	-0.07	0.02
Implicature * Age	0.09	0.03	0.04	0.14
More features * Age	0.02	0.02	-0.02	0.07
Control-double * More features	0.09	0.04	0.02	0.17
Implicature * More features	-0.04	0.06	-0.15	0.07
Control-double * Age * More features	0.01	0.03	-0.05	0.06
Implicature * Age * More features	-0.07	0.04	-0.13	0.00

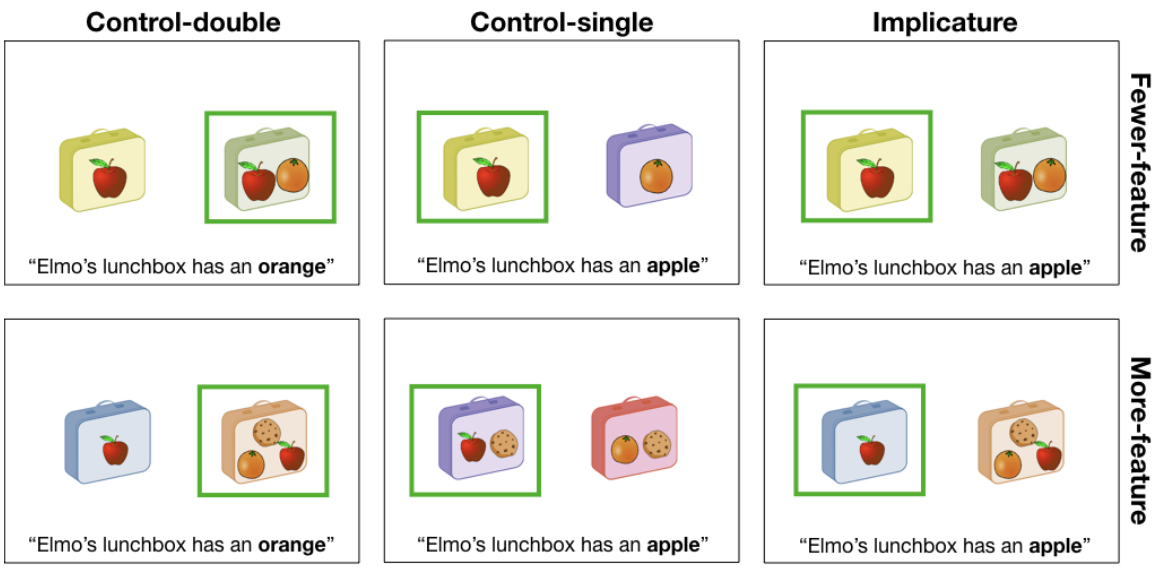

Figure 1. Trial types. Green box indicates the target referent for each trial given the utterance at the bottom.

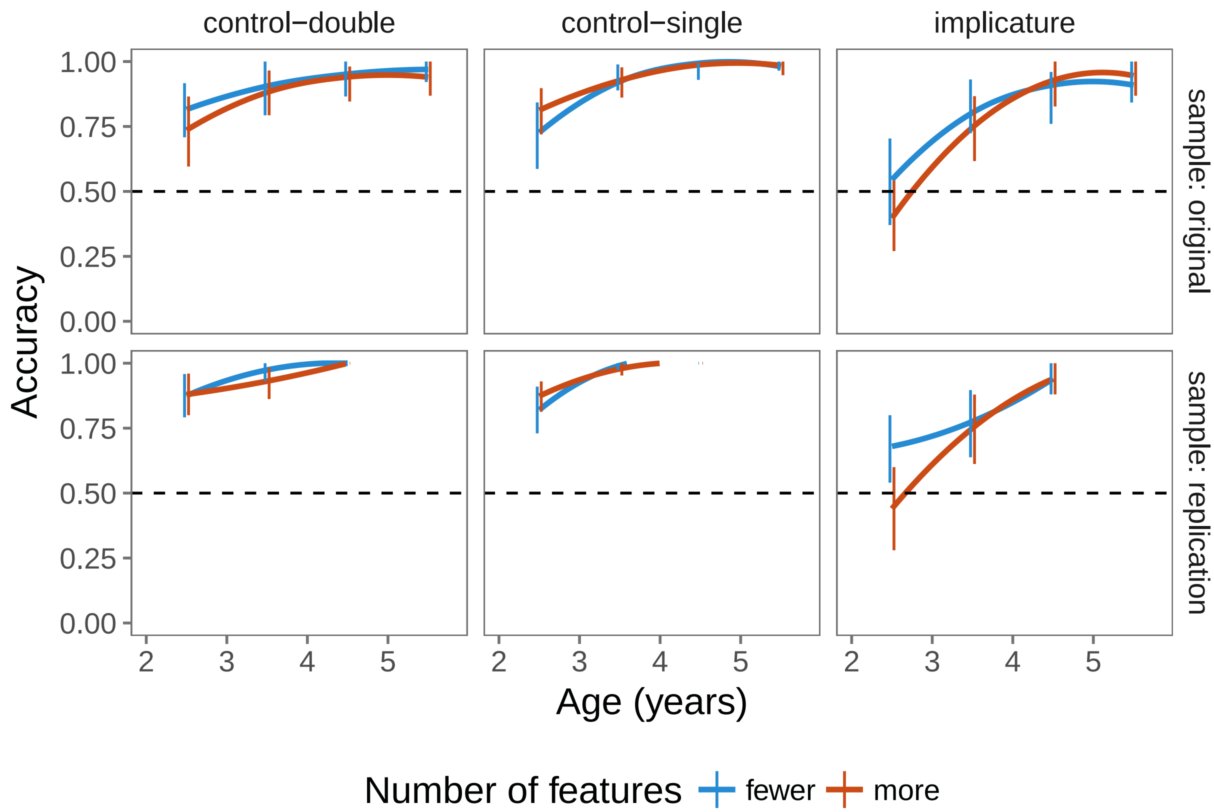

Figure 2. Proportion of 2- to 5-year-old children selecting the target in the original and replication samples (rows) in different trial types (columns). Data are binned into 6-month age groups for visualization purposes (all analyses are conducted on continuous data). Lines are loess smoothing functions. Blue lines represent trials in which there were fewer features present (2-vs-1 for control-double and implicature, 1-vs-1 for control-single) and red lines represent trials with more features (3-vs-1 for control-double and implicature, 2-vs-2 for control-single). Error bars are 95% confidence intervals, and are placed at the mean of the age bin and offset slightly to avoid overplotting. Dashed line represents a conservative chance level at 50%.

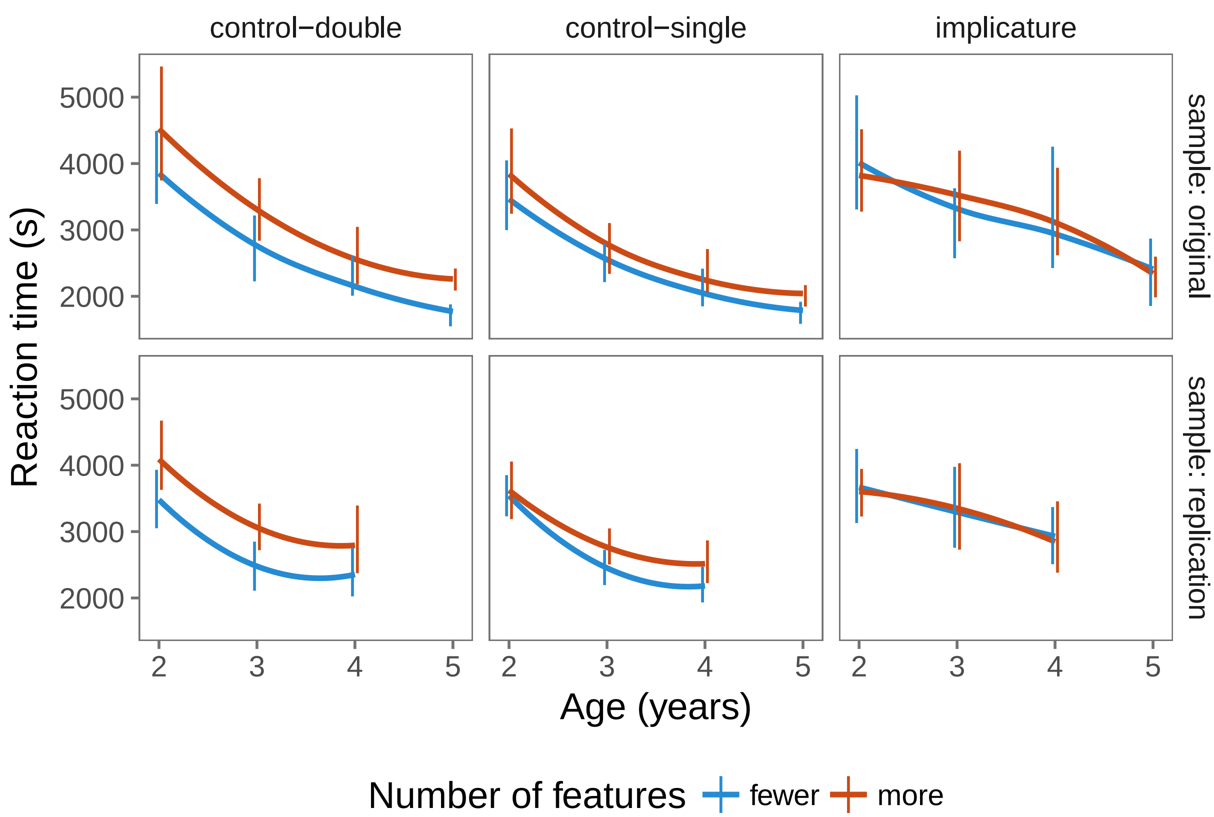

Figure 3. Reaction time to select the correct target referent. Conventions are identical to Figure 2.