Manuscript Details

Manuscript number	JECP_2018_255_R1
Title	The role of salience in young children's processing of ad-hoc implicatures
Article type	Full Length Article

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 supported this hypothesis: Younger children's failures with implicatures were related to effects of the salience mismatch between possible interpretations.

Keywords	Pragmatics; cognitive development; language processing; implicature; tablet
Taxonomy	Language Development, Cognitive Development

Corresponding Author	Erica Yoon
Order of Authors	Erica Yoon, Michael Frank

Submission Files Included in this PDF

File Name [File Type]

cover_letter.docx [Cover Letter]

replies.docx [Response to Reviewers (without Author Details)]

Highlights.docx [Highlights]

abstract.docx [Abstract]

title_page.docx [Title Page (with Author Details)]

manuscript.pdf [Manuscript (without Author Details)]

Submission Files Not Included in this PDF

File Name [File Type]

LaTeX source files.zip [LaTeX Source File]

To view all the submission files, including those not included in the PDF, click on the manuscript title on your EVISE Homepage, then click 'Download zip file'.

Research Data Related to this Submission

Data set https://github.com/ejyoon/simpimp_rs

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

experiment, data, model and manuscript for the project "The role of salience in young children's processing of ad-hoc implicatures"

Department of Psychology
Stanford University
Building 420 (Jordan Hall)
450 Serra Mall
Stanford, CA 94305

650-924-5675
ejyoon@stanford.edu

December 28, 2018

Dr. Catherine A. Haden
Associate Editor, Journal of Experimental Child Psychology

Dear Dr. Haden,

Thank you very much for your comments on our submission to Journal of Experimental Child Psychology, "The role of salience in young children's processing of ad-hoc implicatures."

We have revised our manuscript, focusing especially on improving the discussion of the theoretical contribution based on reviewers' helpful suggestions. In the submitted revision, you will find that the introduction has been almost entirely rewritten to focus on the way that our current experiments are derived from the theoretical literature. Also, we have evaluated our manuscript according to the revision checklist you have provided.

We appreciate your consideration of our revision; please do not hesitate to contact us with any questions or concerns.

Sincerely,
Erica Yoon and Michael Frank
Stanford University

Replies to reviewers' comments:

Reviewer 1

> *This is an interesting paper that presents valuable new data on the acquisition of the capability to infer ad hoc implicatures. My main concern with the paper as it stands is that it doesn't quite do justice to the theoretical issues around implicature, particularly with respect to the question of whether the participants' choices are ultimately motivated by implicature in the strict sense, or some less categorical inference.*
>
> *This point was already made in the previous round of review, but just to underline it: consider the discussion of Stiller et al. (2015) on p.5. There, children are asked to choose between a face with glasses and one with glasses and a hat, based on the description "My friend has glasses". This certainly could be an ad hoc implicature (with "glasses" interpreted as "glasses but no hat"); indeed, it could even be an exhausitivity inference (with "glasses" interpreted as "glasses but no other distinguishing features"). It could also be a probabilistic best guess - suppose we're equally likely to make reference to the glasses or the hat, and we choose one feature, and we're equally likely to be talking about each face. Then two-thirds of the time we refer to "glasses", we're talking about the face with glasses and no hat, and hence that's the one a rational hearer should (always) choose, as their best guess. (This is perhaps implicit in the turn of phrase the authors later use on that same page, where they talk more carefully about "the implicature-consistent choice", rather than, say, "the implicature-driven choice").*
>
> *This may seem like a rather pedantic objection to raise - who cares whether these inferences are strictly speaking implicatures or not? However, as it stands, the key point of this paper is to argue that (specifically) implicature and salience are competing factors in children's decision-making in this kind of experimental paradigm. It strikes me quite forcibly that rational non-implicature inference could also be in the mix, and would compete with salience too. Moreover, the variation in the number of features of the competitor presents a very natural way to explore precisely this question, so I think it should be discussed.*

The reviewer raises the important point that there are many accounts of implicature present in the theoretical literature. These accounts include standard Gricean and neo-Gricean accounts, grammatical accounts that use an EXH operator to exhaustify possible alternatives, and a variety of game-theoretic and probabilistic accounts that use recursive reasoning (listeners reasoning about speakers and vice versa) to explain Gricean behaviors. This theoretical space is complex, and the current study does not allow us to differentiate between these proposals.

That is not to say that we do not have a position on this topic. Goodman & Frank, 2016, *Trends in Cognitive Sciences* reviews the rational speech act framework for probabilistic computation of pragmatic inferences; Potts et al. (2017), *Journal of Semantics* gives some thoughts about how grammatical and probabilistic views can be integrated. But comparing these viewpoints using data is complex, as the predictions of each model are not always well-specified. For example, we believe it's tricky to differentiate what the reviewer calls an "ad-hoc" vs. "exhaustivity" inference here – in both, some alternatives are negated, but the question for all theories is which alternatives those are! No current theory gives a good account of how such alternatives are specified in these kind of situation-specific inferences.

For the purposes of the current manuscript, however, one point is important. We believe that <u>all of the alternatives on the table are implicatures, thus the claim that we are studying implicature is well motivated</u>. The competitor suggested by the reviewer is a "probabilistic best guess" account. On close inspection, this account looks almost identical to a rational speech act model of the type described in Goodman & Frank (2016) and Frank & Goodman (2012): a rational listener reasons about a speaker's choices given particular communicative aims, then makes a "guess" based on those probabilities. Those models are intended to capture implicature behavior. The idea of a guess – a probabilistic choice – is inherent in linking any probabilistic model to human responses; it should not be construed as differentiating the behavior from what is conventionally meant by the term "implicature." (Whether RSA models capture implicature by making it more continuous with general ambiguity resolution – and hence remove the need for a separate term – is a deeper philosophical issue, of course!)

Based on the reviewer's comment, we have rewritten nearly the entire introduction. While it is longer, we hope the reviewer now will agree that it describes the theoretical space more extensively and more accurately. This rewrite also gave us the opportunity to discuss the relationship of our current experiments to the RSA model.

> *There are several other areas where I felt that the paper could be improved, generally because of slightly loose appeals to theory: I note these below.*

We apologize for the theoretical looseness – as noted above, we have tried to avoid making unwarranted theoretical claims, but we do believe that our work here should constrain theorizing about implicature development. We have tried to clarify the direct connections to the RSA model as well as caveating some of the other theoretical claims.

> *Lines 81-83: "In this task, there was no need for children to spontaneously generate the alternative ("glasses and hat")...because the alternative was visible in the context." That isn't really true: "the alternative" is a linguistic expression, and what is visible in the*

context are some objects. So there seems to be quite a strong assumption about preferred encoding lurking within this remark.

Thanks, we now have rewritten that sentence. The revised text reads:

“In this task, the alternative referent (face with glasses and hat) was visible in the context, and thus access to the alternative terms ("glasses” and “hat") was made easier. In general, we assume that the standard route for referring to these visual properties of the context will be by naming them. The design intention in this study for using simple nouns like “hat” was therefore to make it obvious what the linguistic alternatives would be by virtue of the highly accessible names for stimuli.”

Lines 112-114: "Further...what referent a speaker is talking about." I found this point a bit elusive: could it be made more concrete?

Thanks, we have now expanded this point with a more concrete example and explanation.

p.7, footnote 2: Quite a bit of argumentation and a couple of conclusions seem to be brought into this footnote - I think they'd be better later in the main text.

Thank you, we have now moved the content of this footnote to later in the text, in the Discussion section.

Line 205: "due to the strengthening of implicatures". Could you be clearer what you mean by this?

Yes, thanks, we now have modified the wording of this part based on feedback from all three reviewers:

“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 – "Elmo's lunchbox has an apple" is more likely to mean "apple only" given an orange AND cookie on the alternative referent, thus more things that could have been named but were not. On the contrary, younger children were predicted to choose the target less often in the more-feature trials than the fewer-feature trials due to increased saliency of the distractor.”

Lines 274-276: "indicating that the speed of implicature computation did not improve with age as much as the speed of processing unambiguous meanings". I'm not clear on how your analysis distinguishes between implicature computation and the guessing of implicature-compatible answers (let alone the other possibilities I'm arguing for). It *seems a bit premature to conclude anything about implicature computation based on data that must uncontroversially include some points that don't involve implicature and were contributed by participants who were not competent with implicature. (Presumably, if three-year-olds are getting, say, 70% rates of implicature-compatible responses, in a binary choice task, that would suggest that crudely speaking 40% of them are figuring it out correctly and 60% of them are guessing. In that case, 3/7 of the correct responses are just guesses, the timing of which doesn't say anything about implicature on anyone's analysis.)*

As a starting point to this comment, we want to note that our interpretation of reaction time data is very limited. We write that "speed of implicature computation did not improve with age as much as the speed of processing unambiguous meanings." We now revise this to read "reaction time on implicature trials" rather than "implicature computation."

More broadly, we agree that choice reaction time is complex to interpret; in the case of true uncertainty about whether the subject has a particular piece of knowledge, it's not correct just to say that the mean RT is the speed of the computation. That said, we do not think that the two are unconnected, and we hope we can clarify our reasoning.

First, there are two different "guessing" mechanisms being discussed. The first is the "probabilistic guess" account described above. In our response, we hope we have convinced the reviewer that such an account is in fact an account of implicature under a model like the Rational Speech Act model; indeed, this is the account we favor. The second kind of guess, which we will treat here, is a "random guess." This kind of guess reflects an uninformed choice – think, pressing a side randomly, without listening – rather than a probabilistic judgment on the basis of some uncertainty. In any psychological task, there will be some proportion of trials on which participants guess randomly (even if this proportion is very low, for example, with adults in an easy task).

But in fact, most psychologists don't think that "random guessing" is why participants get trials wrong. Instead,when participants get the "wrong answer" on a task, they are trying to answer correctly but failing. Modern decision-theoretic models, like the drift diffusion model (e.g., Ratcliff et al., 1999), quantify the tradeoff between speedy responding on a task and accurate responding – the faster you do a complex task, the more you will fail even though you are not trying to guess. (Then there is some small proportion of cases where people are actually failing to do the task; some models add a parameter to deal with this). Although in some of our lab's work we have attempted to use these models with children's data on implicature tasks (e.g., Schneider et al., 2016), this approach raises a complex set of issues and we don't believe it's appropriate here. Fitting such models requires a lot of data and is really at or beyond the state of the art for developmental data; there are only a small handful of papers using DDM with children.

So what can we conclude by looking at reaction time data in this task? We can see the average time to reach a decision – correct or incorrect. If children are *both* faster *and* more accurate in a task (as happens with development in familiar word recognition in our study and many others), it seems clear that they are getting better at that task and there is no speed-accuracy tradeoff. For our implicature condition, we also see that they are getting faster and more accurate, just at a slower rate. If we were seeing "fast guessers" becoming "slow implicature computers" we would predict a speed accuracy tradeoff, which we don't see.

In sum, we hope the reviewer agrees that our findings are consistent with the modest claim we make, that is: developmental improvement in the implicature trials.

> *Lines 279-291: I think the attempt to understand the processes undertaken by the participants is welcome, but the discussion in this paragraph struck me as a bit too speculative.*

Thanks, we agree with your opinion – to make sure it is clearly conveyed to the reader that this possibility is only speculative, we now have added: "The potential advantage of identifying a feature when it is by itself is only speculative, however, and should be examined further in future work."

Reviewer 2

> *I recommend that this paper be published with minor revisions. (Note, although this is a new manuscript, I have commented on these experiments before when they were a part of a larger paper. Most of the concerns I had about the previous draft with respect to these experiments have been addressed in this new, shorter version.) There are only two contentful comments I have.*

Thanks, we are glad we were able to address most of your concerns for the previous manuscript. We are grateful for the reviewer's thoughtful commentary that led to the current, improved version.

> *A) The authors should not talk about informational strength with respect to lexical items. Only propositions convey information and both "some" and "all" (for example) can be used in both stronger or weaker propositions. So, for example on line 33, the authors say*

"all, which is more informative than the alternative some". "Some" is not more informative than "all". Information is communicated by propositions, not words. "John ate all of the cookies" is more informative than "John ate some of the cookies" but "Everyone who ate some of the candy in their hallowe'en bag was sick" is more informative than "Everyone who ate all of the candy in their hallowe'en bag was sick". Sometimes "all" is used in more informative propositions and sometimes in less informative propositions. There is no inherent information value to the word independent of its use to communicate a proposition. The same mistake is made on line 323.

Thanks, we have now fixed these errors.

B) On lines 203 to 207, the authors write "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." I think this is awkwardly worded and, at best, ambiguous. It makes it sound like the authors predict that the older children will have the opposite influence in terms of object salience (i.e., that the more salient the non-target is, the less likely it will be chosen). I don't think this is what the authors intend to say. (If it is, then I don't see how this prediction would follow.) What I think the authors intend to say is that older children will behave in a way that is more consistent with computing the strengthened meaning, whereas younger children will be more likely to be distracted by choosing the object that is more salient but still consistent with the literal meaning. These are what the results demonstrate. I think these lines should be re-written to more accurately reflect the authors' prediction.

Thanks very much for this helpful point, we have now revised the wording, as discussed in our response to R1.

C) In lines 317 to 324, the authors suggest that their salience hypothesis could extend to account for children's behaviour in other tasks, such as felicity judgment tasks and truth value judgment tasks. They claim that it remains "an open empirical question whether the salience mismatch account might explain children's difficulty with these other cases of implicature as well." However, given the nature of the explanation, I don't see how this is possible. In the current task, the children have to choose between two pictures. In the critical trial, one picture is less salient but matches the pragmatically enriched meaning of the sentence whereas the other is more salient but only matches the literal meaning of the sentence. The idea is that children get distracted by the more salient object and choose it instead of choosing the less salient one that matches the enriched meaning. In the other tasks, there is no choice between objects. There is just one situation and *children have to say whether the sentence is true/silly in that situation, or they have to say whether the speaker who uttered the sentence is right/silly. There is no competition between two different scenarios. Object salience is not a possible complication. The authors suggest that "salience" factors might apply to the choice of interpretation (enriched versus not enriched) or the "salience" of alternatives, however no plausible explanation is given of how this could be a factor. Either the authors need to spell out a plausible influence in detail or they need to admit that object-salience is a confounding factor in this task only. I suspect the later option is the only feasible one. I cannot think of how object salience could play any critical role in the other types of tasks that are used to assess children's abilities to compute implicatures.*

Thanks for this important point to consider. While we do agree that not all tasks of implicature computation involve referent selection and thus there can be less influence of salience of alternatives as we address here, we do think that the salience bias may play an important role in other experimental tasks that use children's judgments between possible answers with different degrees of perceptual (visual) or conceptual bias. Our revision highlights the way that the RSA model analyzes perceptual salience as contributing to a prior probability distribution over referents (and hence suggests the analogy to other pragmatic tasks).

Outside of this theoretical connection, there are some further reasons to think that this account may apply more broadly. First, there are other works that have looked at a similar object reference to look at children's processing of scalar implicature (e.g., Horowitz, Schneider & Frank, 2018) where you see some evidence of similar biases toward more salient alternatives (e.g., a picture where "all of the pictures are cats" as opposed to "some"). Second, there also can be competition between verbal answers they could give (e.g., true/right vs. silly). There are some previous works that talk about potential biases children have toward one verbal answer compared to another (though not in implicature literature -- see below, for yes bias to yes-no questions). Indeed, we see similar biases toward incorrect "yes" responses (compared to fewer incorrect "no"s) in Papafragou and Musolino (2003) and Barner et al. (2011) even in conditions where children computed literal/pragmatic meanings relatively well. We now have added a paragraph with this last point.

D) The statement in lines 377 to 379 is too strong. It states that "that younger children's failures with implicatures are likely related to effects of the salience mismatch between possible referents". I think this hold for the specific task that the article is talking about but it cannot hold for children's failures with respect to other types of tasks. The authors should temper this claim by saying that "younger children's failures with implicatures on an referent-choosing task are confounded by the salience mismatch between possible referents". This is the conclusion that the evidence from their experiments supports. Anything stronger is beyond the scope of their paper.

We agree that the generic, present tense wording of this sentence was too strong. We've refocused the sentence on our current findings and the interpretation of this particular context manipulation.

Reviewer 3

> *This is an interesting study on children's ability to compute ad hoc scalar implicatures between the ages of 2 and 5. I have seen an earlier version of this work and the current manuscript is empirically richer and more focused. The study itself is wonderfully done and clearly presented. It is also theoretically more precise, even though some concerns remain.*

Thanks, we are happy to know that the current manuscript is improved from the previous version.

> *The main contribution of the paper continues to be a replication of prior findings on children's understanding of ad hoc implicatures, and evidence for the idea that implicature calculation drops when the distractor is more 'salient' (i.e., visually busier). This finding has some value in a rapidly developing literature on pragmatics but is still limited in at least two ways: (a) It only applies to methodologies that involve a forced choice between alternative visual scenes, and (b) it relies on the assumption that choice of scene means implicature calculation (and not, for instance, the likelihood of choosing a word such as 'apple' to refer to each of the two scenes in Fig. 1). Both of these concerns were raised during the review of the previous version of the paper. The present version has tried to address the first concern (see Introduction) but the text remains vague about how results can be generalized to other paradigms. (In the Conclusions, the paper attempts a connection to the mutual exclusivity literature that has also used visually contrasting scenes as stimuli. But I found the connection very indirect – there are multiple other differences between assigning novel word meanings and calculating implicatures that may explain early success with exclusivity that do not characterize implicature derivation.)*

Thanks for these thoughtful points. With respect to the first, in our revision we have attempted to be as clear as we can about what we see as the connection between the current paradigm and other paradigms. We have to be up front, however, about the fact that this connection is speculative and based partially in our particular theoretical framework (the Rational Speech Act model, see our response to Reviewer 1). Therefore, we have undertaken a fairly significant rewrite of the introduction to highlight these connections.

Despite this, we emphasize that reference resolution is in fact an important – perhaps even the default – mode of language use for young children. The use of pragmatic inferences to resolve uncertainty between referents is likely very common, whereas the frequency of true scalar implicatures is unknown in child-directed speech (and it is known to be relatively low even in adult-directed speech; Degen, 2015, *Semantics and Pragmatics*).

A further benefit of the salience account is that it raises important methodological considerations, not only for referent-selection paradigms (with *visually* salient distractors) but any paradigms where participants might have biases toward more *conceptually* salient answers.
Bias due to salience seems to exist even in verbal answers to questions, e.g., younger children (2-year-olds) show a bias toward yes compared to no in answering verbal questions but this bias goes away with age (Fritzley & Lee, 2003), and there is some evidence that this bias is related to both their verbal ability and inhibitory control (Morigushi, Okanda, & Itakura, 2008). Thus this bias toward perceptually or conceptually salient answers need to be taken into account in designing tasks for younger children, not only for implicature computation but for any domains that make use of children's judgments that are potentially biased between possible answers. We have now added these points to Discussion.

> *The second concern about the linking assumptions between theory and tablet data has not been addressed in this version.*

As we discussed in our response to Reviewer 1, we view choice behavior as reflecting some underlying preference for the implicature-consistent referent. While our task does not differentiate between different varieties of implicature, we believe that alternative accounts all fall into the category of alternative implicature accounts. For example, we do not understand how the reviewer intends that children compute "the likelihood of choosing a word such as 'apple' to refer to each of the two scenes in Fig. 1" without invoking some variant of a Rational Speech Act model, a probabilistic model of pragmatic inference. We hope that the rewrite of the introduction gives a much more thorough treatment of these issues.

> *Furthermore, the paper could do a better job of stating and motivating its hypotheses. The introduction does not state clearly what the main goal of the paper is (see p.7), or the predictions. When predictions do appear, they are not motivated in their entirety. For instance, it is hypothesized (p.10) 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." The origin of the first hypothesis is unclear: what is 'strengthening of implicatures'? Is this simply an intuition or is the hypothesis based on a theoretical assumption? Only the prediction concerning younger children has been mentioned so far so more justification of the prediction concerning older learners is needed. Notice that the data do not offer evidence for this last hypothesis (even though this may be due to a ceiling effect).

Thanks, we agree with this point. In our rewrite of the introduction, we have introduced our predictions much sooner and in much more depth. We think a lot of the issue here came from the fact that we were reasoning from the Rational Speech Act framework but did not clearly say so.

Minor points:
Fn.1: "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": the notion of entailment is independent of discourse context.

Thanks for pointing this out. We now changed the wording to: "For our purposes, this relation is still ad-hoc in the sense that there is no reason for "the cookies" to implicate "the chocolate chip cookies" in discourse contexts in which no chocolate chip cookies are part of common ground."

P.4: "adults robustly compute implicatures, albeit sometimes more slowly than unambiguous literal meanings": the second claim here is actually debated (see more recent papers by Tanenhaus and Grodner and response from Huang and Snedeker in Cognition) so the text needs to be more nuanced.

Thanks, we now say: "... though their processing time can vary depending on the context", and we have added the references you mention.

p.15 and earlier: "by 2 years of age, children begin to be aware that informativeness is important to communication". I agree with the summary of the cited papers – but there is also a long line of papers reporting massive failures to provide informative descriptions in referential communication tasks (unless children are asked for clarification or given feedback after an initial unsuccessful attempt; see work by Matthews, Tomasello, Katsos etc.). It's a puzzle why pointing leads to early successes but verbal communication is often not as informationally rich.

Thanks very much for this point. We think this point about young children's difficulty in producing informative utterances actually underscores children's challenge to have access to relevant information in comprehension and production -- both speakers' need to *be* more informative, and exactly *how* speakers can be more informative. Our study makes both of these things accessible in the context. We now remark on this issue in the fourth paragraph of the Discussion section, as this issue is relevant to implications from both Barner et al. (2011) and our study.

Highlights

- We examined developmental gains in the speed and accuracy of implicature processing.
- 3- to 5-year-olds can compute ad-hoc implicatures, more accurately with age.
- 2-year-olds' difficulty seems related to the salience of the alternative interpretation.

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 supported this hypothesis: Younger children's failures with implicatures were related to effects of the salience mismatch between possible interpretations.

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

Word count: 9136

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

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 supported this hypothesis: Younger children's failures with implicatures were related to effects of the salience mismatch between possible interpretations.

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

Word count: 9136

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

Language comprehension in context often requires inferring an intended meaning that goes beyond the literal semantics of what a speaker says. Consider a speaker who asserts that:

(1) I ate some of the cookies.

A reasonable listener could assume from this sentence that the speaker ate some *but not all of the cookies*. Inferences like this one, known as *implicatures*, are commonplace in conversation and provide one important tool for speakers to use language flexibly. They also are related to a broader set of pragmatic phenomena like underspecification (Levinson, 2000) and politeness (P. Brown & Levinson, 1987).

How does the ability to make pragmatic inferences develop in childhood? A general finding is that implicature follows a relatively delayed trajectory, with even school-aged children sometimes struggling with implicature tasks (Noveck, 2001). A rich literature has explored both these developmental changes and possible hypotheses about the sources of difficulty for children (e.g., Barner, Brooks, & Bale, 2011; Papafragou & Musolino, 2003; Stiller, Goodman, & Frank, 2015). These investigations are important because they shed light on developmental changes in children's ability to comprehend language in context more broadly, as well as the processing challenges posed by pragmatic language comprehension.

In the current paper, we investigate the developmental trajectory of the processing of one specific type of implicatures, *ad-hoc implicatures*, which tend to be easier for young children than other implicatures that rely on more sophisticated linguistic knowledge (Horowitz, Schneider, & Frank, 2018; Papafragou & Musolino, 2003; Stiller et al., 2015). In addition, we test a specific proposal for why young children might find even these inferences challenging, namely that the inferential target is typically less salient than the distractor. In the remainder of the Introduction, before describing our own work we first introduce pragmatic implicature in more depth, then review developmental evidence on implicature.

Pragmatic Implicature

In Grice (1975)'s classic account of pragmatic inference, 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. The listener then expects a cooperative speaker to have produced an utterance that is truthful, informative, relevant, and concise, relative to the the present conversational needs. Based on these expectations, 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.

A concrete example of such an implicature follows from sentence (1), which implicates that the speaker ate some *but not all of the cookies.* This kind of inference is often referred to as a *scalar implicature* because it relies on the fact that "all of the cookies" entails "some of the cookies" as part of a lexical scale (Horn, 1972). In contrast, another kind of implicature, *ad-hoc implicature*, is context-based. Uttering:

(2) 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.*[1] In this case, the context sets up a contrast between the proposition offered ("ate the chocolate chip cookies") and a stronger set of alternatives ("ate [all/both the chocolate chip and the raisin] the cookies") that is determined by the context (and hence is "ad hoc" in the sense of being constructed in this particular situation).

Implicatures like these have been an important case study for pragmatics more broadly. Notably, different accounts of pragmatic reasoning analyze even the simple examples above in different ways. In the classic Gricean analysis (as elaborated by Levinson, 1983), the speaker

[1] Grice (1975) calls these implicatures generalized (scalar) vs. particularized (ad-hoc), but we use a theory-neutral designation here.

utters p ("some of the cookies"), which implicates q ("not all of the cookies") in the following way. (A) The speaker is presumed to be cooperative and observing Grice's maxims. (B) To maintain this assumption, the listener must assume that q is true; otherwise a maxim will be violated. (In this case the maxim is informativeness, since saying "some of the cookies" if "all of the cookies" were true would be underinformative). (C) The speaker is presumed to believe that it is mutually known by both parties that the listener can work out q.

This analysis – though influential – is in fact just one proposal among many, and likely does not map onto either the mental computations carried out by listeners or the specific issues that lead to developmental differences in implicature ability. 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 in different ways. For example, Chierchia, Fox, and Spector (2012) give an account of implicature as a specific, grammaticalized operation that involves enriching the meaning of p with the negation of all stronger alternatives within a specified alternative set. In contrast, on the probabilistic view, implicatures arise naturally as part of the process of cooperative reasoning by rational agents.

Our goal here is not to distinguish between these different formalisms; instead, we are interested in understanding the processing of implicature in childhood. Despite that, it is useful to review the probabilistic view as it helps guide some of our predictions below. We consider sentence (2), following the analysis given in Goodman and Frank (2016). Under the rational speech act (RSA) model, there is a space of meanings (e.g., ATE(chocolate chip & raisin), ATE(chocolate chip), etc.), each of which may have some prior probability of being correct. There is also a space of utterances (e.g., "I ate the chocolate chip cookies," "I ate the cookies"), each of which is either literally consistent or inconsistent with each meaning. Given a particular utterance, a listener can reason probabilistically about the speaker's intended meaning in making this utterance. He can do this by considering that the speaker is a Bayesian agent who chose the appropriate utterance for her intended meaning. He reasons about the speaker making her own choice by considering a listener who is also a Bayesian agent reasoning in this same way. This definition is endlessly recursive, however. In practice, the recursion can be grounded by a speaker considering a "literal listener," who interprets utterances according to their literal truth value (for further formal details, see Goodman & Frank, 2016).

In the specific case of (2), the listener's reasoning can be glossed as "if the speaker had wanted to say she ate 'all of the cookies', she could have said just 'cookies'; but she didn't, she said something more specific: 'chocolate chip'; thus she probably intended me to recover the meaning ATE(chocolate chip)." Notice that this reasoning, when explained verbally, actually approximates the standard Gricean logic (though with some differences). Of course, one benefit of the RSA formalism is that probabilities can be put to each of these inferences and so the strength of the interpretive judgment can be predicted (Frank & Goodman, 2012).

The Development of Pragmatic Implicature

A rich psycholinguistic literature has measured adults' processing of implicatures relative to literal interpretations and has found that adults robustly compute implicatures in a range of contexts, though their processing time can vary depending on the context (Bott, Bailey, & Grodner, 2012; Breheny, Ferguson, & Katsos, 2013; Grodner, Klein, Carbary, & Tanenhaus, 2010; Huang & Snedeker, 2018). 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 on a variety of implicatures.

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, 2009; 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 et al. (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 et al., 2018; Katsos & Bishop, 2011; Papafragou & Tantalou, 2004; Stiller et al., 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, the alternative referent (face with glasses and hat) was visible in the context, and thus access to the alternative terms ("glasses and hat") was made easier. In general, we assume that the standard route for referring to these visual properties of the context will be by naming them. The design intention in this study for using simple nouns like "hat" was therefore to make it obvious what the linguistic alternatives would be by virtue of the highly accessible names for stimuli. 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., 2018).

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 these easier ad-hoc implicature tasks specifically?

The Current Study

One potential explanation for younger children's struggle with ad-hoc implicatures is the mismatch in salience between potential interpretations. This explanation is inspired by the RSA framework described above, in the sense that this salience mismatch would be manifest in the pragmatic computation as a higher prior probability of a particular referent.

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. Under the RSA framework – and very likely under other pragmatic theories, though perhaps with a less clearly specified prediction – such a mismatch in prior probabilities would lead to a weaker pragmatic inference.

The mismatch between stimulus salience (prior probability) and the target of the pragmatic inference may be particularly difficult developmentally. From a mechanistic perspective, a task with this kind of competition between targets may be especially challenging to children because their executive function is not yet fully developed (Davidson, Amso, Anderson, & Diamond, 2006; Diamond & Taylor, 1996), and 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).

Such an issue might be important outside of the specific case of ad-hoc implicature and referent-selection tasks. For example, referent selection tasks may be representative of analogous problems in naturalistic language comprehension for children, in which the goal is often to figure out what referent a speaker is talking about or how to connect a new word to a new referent (in the case of word learning; Frank & Goodman, 2014). And in such situations, there is a body of evidence suggesting that referent salience does in fact influence children's attention (Hollich et al., 2000; Yurovsky & Frank, 2017).

Further, under RSA analysis given above there is no fundamental difference between referent selection tasks and other implicature comprehension tasks. Thus, the asymmetry between correct but weaker target meaning and incorrect but more salient or higher prior probability distractor meaning is present in other types of implicatures too, though less obviously so. For example, scalar implicature is typically described as rejecting the term that yields the "stronger" propositional meaning (e.g., ate "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. 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, 2009). At least in these cases – and perhaps more generally – the stronger alternative could reasonably be viewed as being more salient or higher prior probability. And, as above, when prior probabilities (whether induced by perceptual factors like salience or emerging from other sources) conflict with pragmatic inferences, the resulting comprehension situation may be especially difficult for children.

For all of these reasons, in our current work, we were interested in exploring the issue of distractor salience and how it played out in the development of implicature processing for children. 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., 2018; Stiller et al., 2015), and are analogous to one important aspect of of language comprehension in naturalistic language environments, namely identifying a speaker's intended referent.

This setup allowed us to create a systematic manipulation of the stimuli in our referent selection method. Under the RSA model, the more alternative utterances there are to refer to a particular referent, the less likely any one of them is. Thus, adding more features to the distractor referent in the referent selection task should make it even less likely as the referent of any particular one. For example, in the faces case used by Stiller et al. (2015), if the target is a face with glasses, then a face with a hat, glasses, and a mustache (three features) should be a *worse* distractor referent for "glasses" than a face with just a hat and glasses (two features). Frank and Goodman (2014) tested this prediction with adults in a word-learning case and found quantitative support for the idea that the number of features was related to the strength of pragmatic judgments.

The interesting thing about this manipulation, however, is that it might very well have an *opposite* effect on young children because of the referent salience explanation given above. While a distractor with more features should create a stronger pragmatic inference, it should also be more salient to young children, leading to a higher prior probability and worse performance. Thus, in our current experiment we predicted that young children would struggle differentially in the case there were more features on the distractor, while older children would find this case no more difficult and perhaps even easier.

We stress that, although our manipulation was inspired by the RSA model, it does not depend on that model. As touched on above, there are a variety of different accounts that try to explain exactly what pragmatic inference children are making in ad-hoc implicature tasks. In the Stiller et al. (2015) example, "my friend has glasses" can implicate "my friend has glasses *but no hat*" based on the immediate context. A slightly different interpretation could be: "…glasses *but no other distinguishing features*," however (exhaustivity implicature; Groenendijk & Stokhof, 1985). For the purposes of the current work, we cannot differentiate between these proposals – as long as an account incorporated prior information in some fashion, it would likely make a similar prediction. Thus, our goal is not to make a test of a particular implicature account, but rather to test the idea that referent salience (instantiated as prior probability in the RSA model) affects children's implicature behavior.

In our experiment, we implemented the referent selection task using a tablet paradigm. This methodological change allowed us to examine children's reaction times for selecting the target referent along with their specific selection (Frank, Sugarman, Horowitz, Lewis, & Yurovsky, 2016). Compared to previous studies, we also 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 the younger children in our task.

We present data here 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 across the three youngest groups (2-, 3- and 4-year-olds) to replicate this initial finding and make a stronger test of the hypothesis.

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 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 because implicatures are strengthened more in more-feature trials – "Elmo's lunchbox has an apple" is more likely to mean "apple only" given an orange AND cookie on the alternative referent, thus more things that could have been named but were not. On the contrary, younger children were predicted to choose the target less often in the more-feature trials than the fewer-feature trials because the distractor is more salient in the fewer-feature trials, while still being consistent with the literal meaning.

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 0.2.2; Müller, 2017a), *brms* (Version 2.0.1; Bürkner, 2017), *dplyr* (Version 0.7.7; Wickham, Francois, Henry, & Müller, 2017), *forcats* (Version 0.2.0; Wickham, 2017a), *ggplot2* (Version 3.0.0; 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 0.2.5; Henry & Wickham, 2017), *Rcpp* (Eddelbuettel & Balamuta, 2017; Version 0.12.19; Eddelbuettel & François, 2011), *readr* (Version 1.1.1; Wickham, Hester, & Francois, 2017), *stringr* (Version 1.3.1; Wickham, 2017b), *tibble* (Version 1.4.2; 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. replication) 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., 2018; Stiller et al., 2015). In our paradigm, even 3-year-olds chose the inferential target above chance[2] (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

[2] 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).

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). 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. Thus, this result supports our hypothesis that the salience contrast between conditions led to greater difficulty with the implicature task for for younger children.

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 reaction time on implicature trials 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.

There was no interaction between inference trials and number of features, or between inference trial, age and number of features, however. 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. The potential advantage of identifying a feature when it is by itself is only speculative, however, and should be examined further in future work.

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. Further, 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). Inspired by this pattern and following the predictions of the RSA model of pragmatic inference, 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, with more mixed support for the second part.

In particular, although we observed numerical hints of a gain in accuracy for older children in one sample, we did not see a consistent facilitation effect due to our manipulation. We suspect this finding is due to a ceiling effect: Referent selection via ad-hoc implicature is relatively trivial for four-year-olds (see also Horowitz et al., 2018). 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.

Our findings here support the idea that salience-related competition plays a role in young children's difficulties with ad-hoc implicature. Our salience account is most manifest in the kind of simple referent selection tasks we used here. Despite this, following the general mapping of perceptual salience to prior probabilities in the RSA framework more broadly, we speculate that the account may apply 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 a salience mismatch account – perhaps relabeled as a prior probability mismatch – might explain some aspects of children's difficulty with these other cases of implicatures as well.

The salience hypothesis we tested here relates to broader methodological issues in experiments for young children with both visual and verbal alternative responses. One example of such a bias is the tendency of 2-year-olds to show a bias toward "yes" compared to "no" in answering questions. This bias disappears with age (Fritzley & Lee, 2003), and there is some evidence that it is related to children's verbal ability and inhibitory control (Moriguchi, Okanda, & Itakura, 2008). In general, as work on pragmatic inference begins to examine younger children's abilities, it is important to take into consideration a range of cognitive factors in task design.

Following this line of reasoning, one further potential 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 feature, even though in many respects this task should be more, not less, difficult than ad-hoc implicature. One reason for this success might be that 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. We might speculate 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. 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 inferential alternatives is only beginning to emerge in the period from 4 top 6 (Barner et al., 2011; Horowitz et al., 2018; 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).

As illlustrated by this timeline, the salience hypothesis we tested is not mutually exclusive with other accounts of children's difficulties in implicature. For example, the "alternatives hypothesis" (Barner et al., 2011) is independently supported by a variety of experiments (Horowitz et al., 2018; Skordos & Papafragou, 2016). Indeed, both the salience and alternatives hypotheses are likely true and likely contribute to children's difficulty with implicatures to different degrees in different tasks and at different ages.

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., 2018; 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 (and thus greater processing time) might have been caused by a real desire to acquire more information, rather than the mere perceptual salience of the distractors. Second, as noted in the Introduction, our study does not differentiate between different theoretical proposals about how pragmatic inference is being computed in the current task. However, we believe that we are addressing development of implicatures in general. Third, 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 on an referent-choosing task are confounded by 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. J. (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

Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage* (Vol. 4).

Cambridge University Press.

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).

Chierchia, G., Fox, D., & Spector, B. (2012). Scalar implicature as a grammatical phenomenon. In Maienborn, von Heusinger, & Portner (Eds.), *Semantics: An international handbook of natural language meaning* (Vol. 3). New York: Mouton de Gruyter.

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

Fritzley, H. V., & Lee, K. (2003). Do young children always say yes to yes–no questions? A metadevelopmental study of the affirmation bias. *Child Development*, *74*(5), 1297–1313.

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.

Grodner, D. J., Klein, N. M., Carbary, K. M., & Tanenhaus, M. K. (2010). ?Some,? And possibly all, scalar inferences are not delayed: Evidence for immediate pragmatic enrichment. *Cognition*, *116*(1), 42–55.

Groenendijk, J., & Stokhof, M. (1985). On the semantics of questions and the pragmatics of answers. *Semantics: Critical Concepts in Linguistics*, *288*.

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

Hollich, G. J., Hirsh-Pasek, K., Golinkoff, R. M., Brand, R. J., Brown, E., Chung, H. L., … Bloom, L. (2000). Breaking the language barrier: An emergentist coalition model for the origins of word learning. *Monographs of the Society for Research in Child Development*, i–135. doi:10.1111/1540-5834.00090

Horn, L. R. (1972). *On the semantic properties of logical operators in english* (PhD thesis). University of California, Los Angeles.

Horowitz, A. C., Schneider, R. M., & Frank, M. C. (2018). The trouble with quantifiers: Exploring children's deficits in scalar implicature. *Child Development*, *89*(6), e572–e593.

Huang, Y. T., & Snedeker, J. (2009). 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

Huang, Y. T., & Snedeker, J. (2018). Some inferences still take time: Prosody, predictability, and the speed of scalar implicatures. *Cognitive Psychology*, *102*, 105–126.

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

Levinson, S. C. (1983). Pragmatics. cambridge textbooks in linguistics. *Cambridge/New York*.

Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. MIT press.

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.

Moriguchi, Y., Okanda, M., & Itakura, S. (2008). Young children's yes bias: How does it relate to verbal ability, inhibitory control, and theory of mind? *First Language*, *28*(4), 431–442.

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

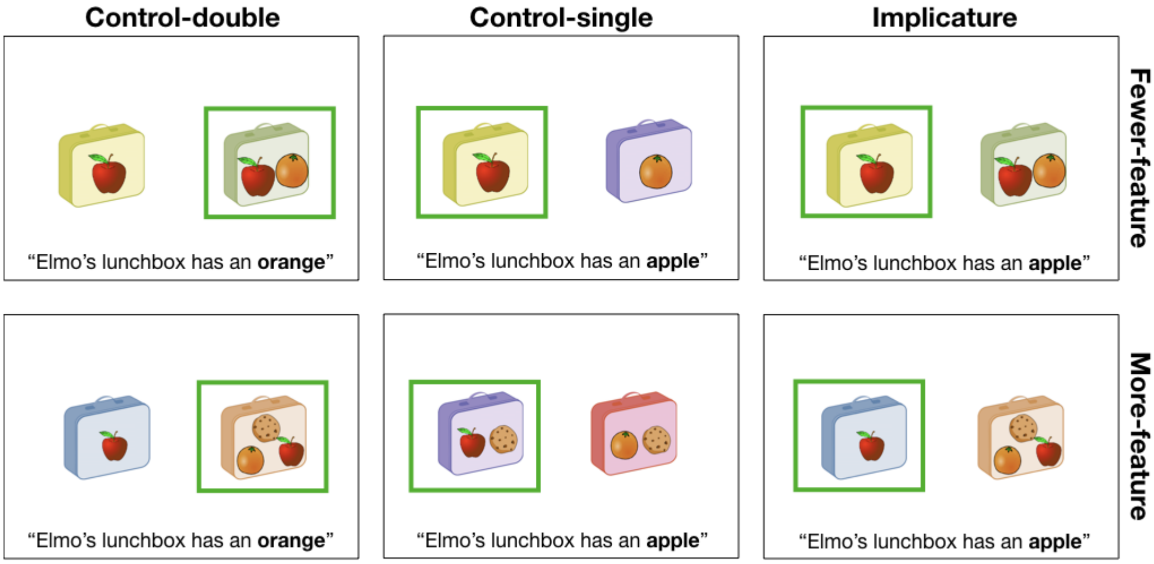

Figure 1. Trial types. Green box indicates the target referent for each trial given the utterance at the bottom.

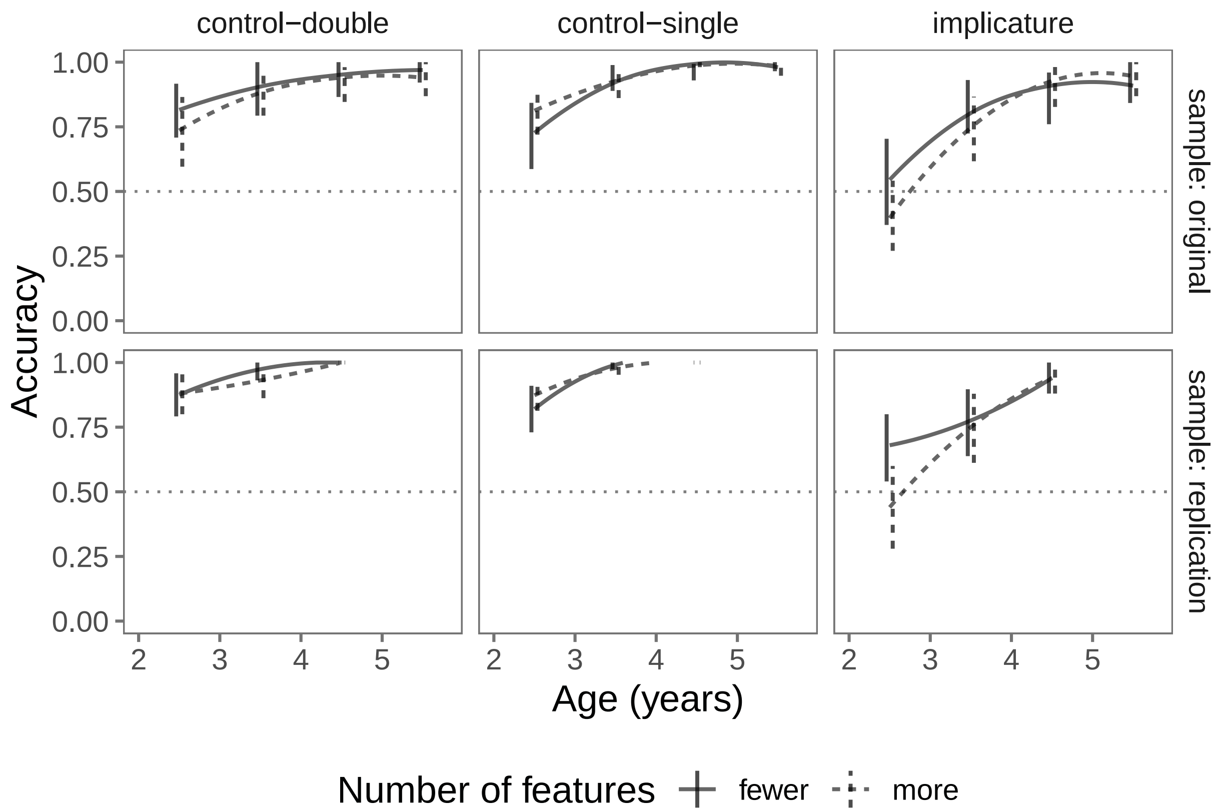

Figure 2. Proportion of 2- to 5-year-old children selecting the target in the original and replication samples (rows) in different trial types (columns). Data are binned into 6-month age groups for visualization purposes (all analyses are conducted on continuous data). Lines are loess smoothing functions. Solid 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 dashed 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. Dotted line represents a conservative chance level at 50%.

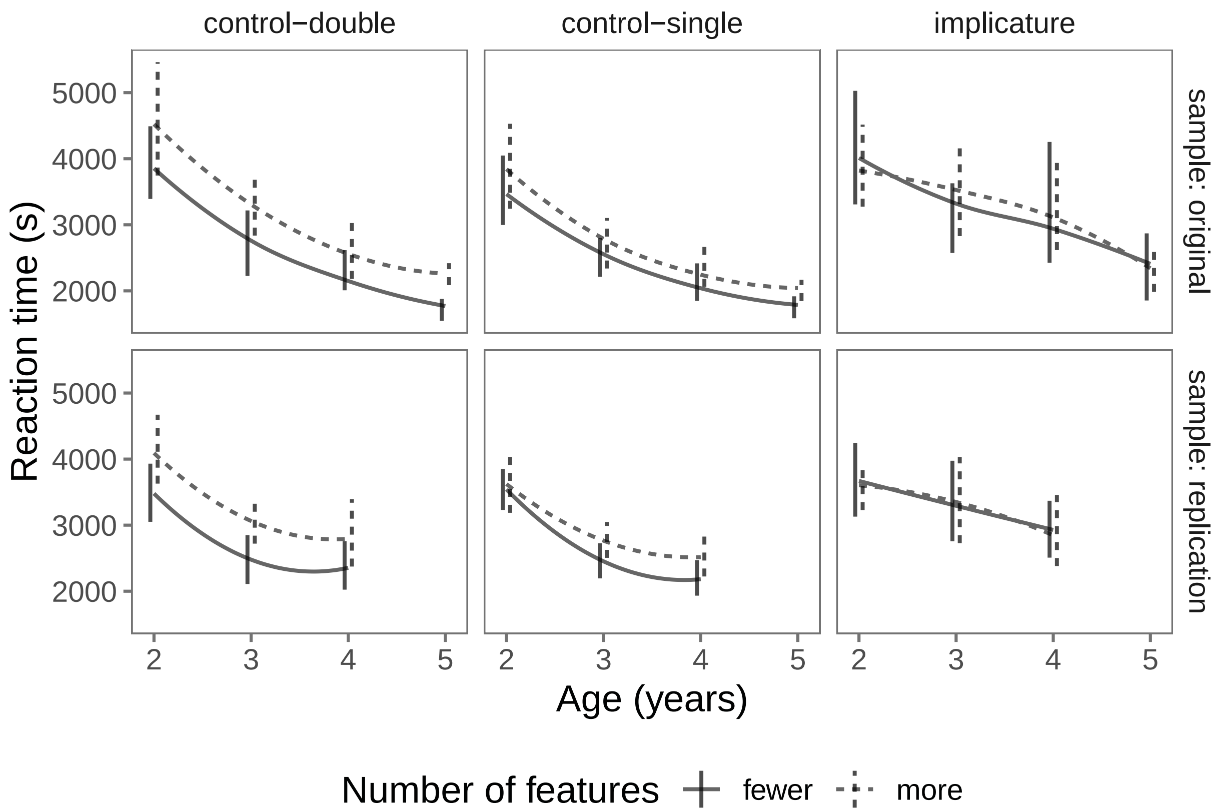

Figure 3. Reaction time to select the correct target referent. Conventions are identical to Figure 2.