Preschoolers flexibly adapt to linguistic input in a noisy channel

Daniel Yurovsky[1], Sarah Case[2], and Michael C. Frank[2]

[1]Department of Psychology, University of Chicago

[2]Department of Psychology, Stanford University

Author Note

Please address correspondence to:

Daniel Yurovsky

Department of Psychology

University of Chicago

5848 S University Avenue

Chicago IL, 60637

Email: yurovsky@uchicago.edu

Abstract

Because linguistic communication is inherently noisy and uncertain, adult language comprehenders integrate bottom-up cues from speech perception with top-down expectations about what speakers are likely to say. Further, in line with the predictions of ideal observer models, comprehenders flexibly adapt how much they rely on these two kinds of cues in proportion to their changing reliability. Do children also show evidence of flexible, expectation-based language comprehension? We presented preschoolers with ambiguous utterances that afforded two different interpretations, depending on whether they privileged perceptual input or top-down expectations. Across three experiments, we manipulated the reliability of both their perceptual input and their expectations about the speaker's intended meaning. As predicted by noisy channel models of speech processing, 4- and 5-year-old—but not younger—children flexibly adjusted their interpretations as cues changed in reliability.

Keywords: Language processing, noisy channel, cognitive development

Preschoolers flexibly adapt to linguistic input in a noisy channel

Imagine Bob hears Alice say "I had carrots and bees for dinner." Perhaps she visited an exotic restaurant, and he should ask how they tasted. Or perhaps he misheard her or she misspoke—she actually ate *peas*. To interpret Alice's utterance, Bob has to integrate perceptual information from her speech with his expectations about what words usually go with "carrots" and "dinner" and what foods people usually eat. Modern statistical language processing systems use a body of theory based on this idea—that language is a *noisy channel*, and that Bob can correct for perceptual errors using linguistic expectations about what Alice was likely trying to say (Jelinek, 1976; Shannon, 1948).

Noisy channel principles provide a powerful framework for explaining how people process language in complex and uncertain real-time communicative situations (Clayards, Tanenhaus, Aslin, & Jacobs, 2008; Levy, 2008; Jaeger, 2010). On this view of language, comprehenders integrate prior expectations with perceptual data probabilistically, weighting each according to its reliability (Ernst & Banks, 2002; Jacobs, 1999). In one demonstration of such integration, Gibson, Bergen, and Piantadosi (2013) presented participants with semantically implausible sentences (e.g., "The mother gave the candle the daughter"), which could have been produced by small typographical errors in otherwise much more plausible sentences ("The mother gave the candle *to* the daughter"). Adults corrected these errors, and critically did so -more often when they thought the communicative channel was noisy (and hence the perceptual signal was unreliable). Conversely, adults corrected these errors less often when they thought they were in a "silly" context where many other sentences were similarly implausible.

Do children also process language in the same flexible, expectation-based way? Toddlers use speakers' social and pragmatic cues to determine their intended referent in otherwise ambiguous situations (Carpenter, Nagell, Tomasello, Butterworth, & Moore, 1998; E. V. Clark, 2009). They also use acoustic cues like speaker identity and linguistic cues like grammatical gender (Lew-Williams & Fernald, 2007; Creel, 2012). However, these successes have been shown only when all cues point to the same meaning. When top-down and bottom-up cues conflict, preschoolers often over-weigh, or even attend exclusively to lower-level cues (Trueswell, Sekerina, Hill, & Logrip, 1999; Snedeker & Trueswell, 2004).

Outside of language processing, even older children sometimes fail to combine cues (Gori, Del Viva, Sandini, & Burr, 2008; Nardini, Jones, Bedford, & Braddick, 2008; Nardini, Bedford, & Mareschal, 2010). For example, while adults integrate haptic and visual cues to estimate both size and orientation, 8-year-olds rely exclusively on vision for orientation and haptics for size. Thus, children's successful use of independent information—for example, about high-level speaker expectations(Graham, Sedivy, & Khu, 2014; Matthews, Lieven, & Tomasello, 2010) or speaker reliability (Pasquini, Corriveau, Koenig, & Harris, 2007)—does not their guarantee integration with perceptual uncertainty. Our experiments test adaptive, expectation-based integration in language in the absence of other processing demands.

We created a paradigm to independently manipulate expectations about speaker plausibility and perceptual noise. We introduced preschoolers (and adults) to either a Plausible or Implausible Speaker who initially uttered unambiguously different sentences like "my cat has three little [kittens/hammers]" (Figure 1a). Participants were then asked to resolve the intended meaning for ambiguous sentences like the "bees/peas" example above, which could either be produced by a perceptual error, or could convey implausible content (Figure 1b). If children integrate speaker expectations and channel noise, their interpretations should be a product of both.

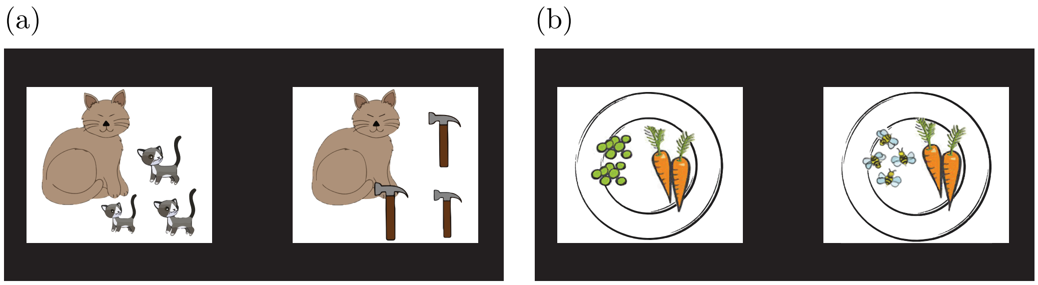

Figure 1. Example Exposure and Test trials. On Exposure trials, the two pictures and their corresponding referring expressions were highly distinct (e.g. "my cat has three little [kittens/hammers]." In contrast, the two pictures on Test trials could be referred to by expressions that contained a single phonological error (e.g. "I had carrots and [bees/peas] for dinner." Children and adults introduced to a speaker who consistently referred to the plausible referent on Exposure trials were more likely to interpret a reference to the implausible picture on Test trials ("bees") as a reference to the plausible picture ("peas").

Experiment 1

Method

Participants. Children were recruited at the Bing Nursery School on Stanford's campus. Each child was asked if they would be willing to play a game with the experimenter, and informed that they could stop playing at any time. Children were randomly assigned to speaker conditions, and we collected data until there were at least 20 participants in each condition (similar to other psycholinguistic studies of children e.g. Trueswell et al., 1999; Creel, 2012). Data from a total of 43 children were collected; children were all between 4 and 6 years old, and approximately half were female. Neither age nor sex distribution varied significantly across conditions (Plausible: 23 children [12 girls], Mean age = 4.6 years [range = 4.0–5.3 years], Implausible: 20 children [10 girls], Mean age = 4.7 years [range = 4.1–5.4 years]).

Adult participants were recruited through Amazon Mechanical Turk. We posted 50

Human Intelligence Tasks (HITs) to be completed only by participants with US IP addresses that each paid 30 cents. A total of 50 HITs were posted, with Speaker condition assigned randomly to each participant. A sample size of 50 was chosen on the basis of the effect size in the child data.

Stimuli, Design, and Procedure. Experiment 1 consisted of a series of trials on which participants saw two pictures and heard a sentence referring to one of them. Pictures were constructed from clipart freely available on the internet, and audio was recorded by a female native speaker of English. In order to increase the ambiguity of the spoken utterances, and thus give us more power to detect error-correction, all of the audio recordings were convolved with Brown noise of amplitude $\sim$.6 using the Audacity audio editor and then played back through computer speakers to produce additional distortion.

Each trial contained one semantically plausible picture and one semantically implausible picture. On Exposure trials, these pictures and their corresponding referring expressions were highly distinct (Figure 1a). In contrast, on Test trials the two pictures corresponded to referring expressions that were only one consonant or vowel apart (Figure 1b). Each participant saw eight Exposure trials, followed by seven Test trials. The order of these trials within each of these two blocks was randomized across participants, as was the on-screen position of the two pictures on each trial. For participants in the Plausible Speaker condition, the speaker referred to the plausible referent on each of the eight Exposure trials. In contrast, for participants in the Implausible Speaker condition, the speaker referred to the implausible referent on each Exposure trial. In both conditions, the speaker referred to the implausible referent on all seven Test trials.

For all participants, the experiment began with a short introduction to Katie, the speaker who would be referring to pictures for the remainder of the task. After seeing her picture and being introduced to her, participants completed the remaining trials, responding either with the mouse (adults), or by touching one of the pictures on an iPad (children). Adults were instructed by a series of written prompts; children were given instructions by a live experimenter. Three children's responses were coded from video due to a software error. All stimuli, data, and analyses are available at `http://github.com/dyurovsky/noisy-kids`.

Results

In order to validate our manipulation, we first analyze the effect of the speaker on Exposure trials. If participants were attending to the speaker's requests during Exposure trials, those in the Plausible condition should have been more likely to choose the plausible referent (e.g. kittens), and those in the Implausible condition should have been more likely to choose the implausible referent (e.g. hammers). We tested this prediction formally by fitting a mixed-effects logistic regression predicting choice on Exposure trials separately for children and adults. As predicted, both children and adults selected the plausible referent more often in the Plausible condition ($\beta_{child} = 4.54$, $z = 8.28$, $p < .001$, $d = 5.56$; $\beta_{adult} = 9.41$, $z = 5.83$, $p < .001$, $d = 24.3$). We then fit a model to all of the data, asking whether adults and children were differentially affected by speaker condition. We found significant effects of condition, age group, and their interaction ($\beta_{Plausible} = 9.52$, $z = 7.87$, $p < .001$, $\beta_{child} = 1.96$, $z = 3.33$, $p < .001$; $\beta_{child\,\cdot\,Plausible} = -5$, $z = -4.12$, $p < .001$). Thus, both children and adults were sensitive to the speaker manipulation during Exposure trials, selecting the appropriate referent whether or not the request was implausible, although adults selected the correct referent more often in both conditions.

Did this exposure to a Plausible vs. Implausible speaker change participants' expectations on the ambiguous Test trials? Figure 2 shows the proportion of both children and adults who selected the plausible referent at test in both conditions. As predicted, both groups were sensitive to the manipulation, selecting the plausible referent more often in the Plausible Speaker condition ($\beta_{child} = 1.10$, $z = 3.53$, $p < .001$, $d = 1.10$; $\beta_{adult} = 3.11$, $z = 3.56$, $p < .001$, $d = 1.09$). While children were overall more likely to pick the plausible referent in both conditions, the effect size of the difference between conditions was nearly

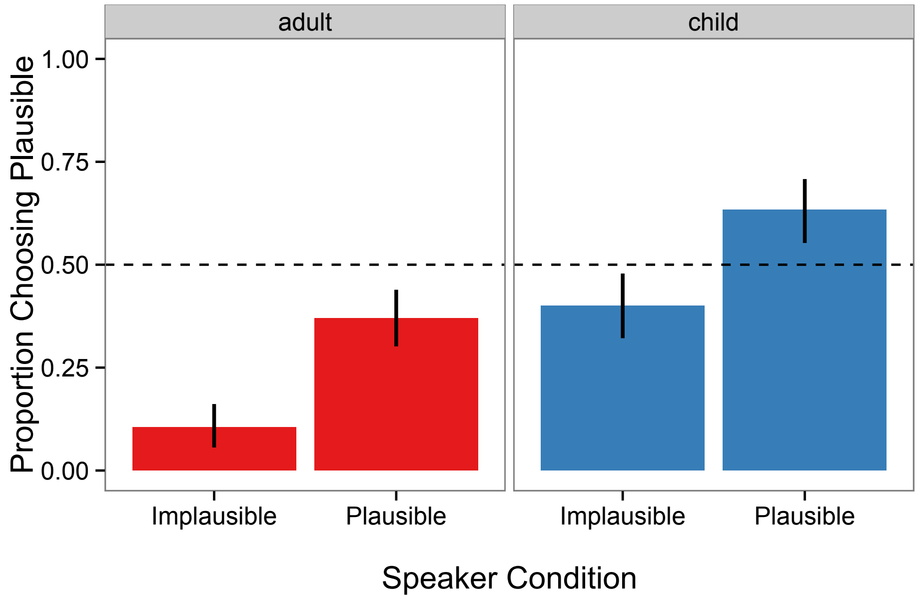

Figure 2. Experiment 1 Test trial selections for both children and adults. In line with the predictions of a noisy channel model of language processing, both children and adults were more likely to correct the phonological error during Test trials when they were previously exposed to a Plausible Speaker. Bars show group-averaged proportion of plausible item selection at Test, error bars show 95% confidence intervals computed by non-parametric bootstrap at the subject level. The dashed line indicates chance performance.

identical in adults and children, indicating equal adaptation to the Plausible vs. Implausible speaker. To confirm these findings formally, we again fit a mixed-effects regression predicting choice on Test trials from group (child vs. adult), condition (Plausible vs. Implausible Speaker), and their interaction. Both of the main effects were significant, but the interaction was not ($\beta_{child} = 1.96$, $z = 3.33$, $p < .001$, $\beta_{Plausible} = 2.33$, $z = 4.40$, $p < .001$, $\beta_{child \cdot Plausible} = -1.01$, $z = -1.53$, $p = .13$). Thus, children and adults, to the same degree, were more likely to select the plausible referent on ambiguous Test trials when the speaker had previously referred to plausible referent on unambiguous Exposure trials.

When children and adults were exposed to a speaker who was likely to produce semantically implausible utterances (e.g. "my cat has three little hammers"), they were more likely to interpret ambiguous utterances literally instead of error-correcting to a more semantically plausible alternative. Intriguingly, the size of this adaptation was nearly identical in both groups, suggesting that 4- and 5-year-olds are already adapting as rapidly as adults. Children were, however, more likely overall to pick the plausible referent during ambiguous test trials, suggesting that they generally rely more on their expectations than do adults. This finding is consonant with other evidence showing significantly more noise in children's perceptual systems (Neuman & Hochberg, 1983).

Experiment 2

Experiment 2 replicates Experiment 1 in a larger and developmentally-broader sample of children, investigating the development of noisy channel processing.

Method

Participants. For Experiment 2, children were recruited from the floor of the San Jose Children's Discovery museum. An experimenter approached the child and parent and obtained informed consent before inviting both to enter a separate room in which an iPad and camera were set up. Data from a total of 146 children were collected, 6 of whom were excluded for $< 50\%$ exposure to English. As before, children were recruited until at least 20 had been run in each condition for each of 3 age groups: 3, 4, and 5-year-olds. Children's ages were comparable across conditions, although gender varied more due to sampling (Table 1).

Stimuli, Design, & Procedure. Experiment 2 was an exact replication of Experiment 1.

Results

As in Experiment 1, we first establish that children understood the task, and responded appropriately to the the Plausible and Implausible speakers on Exposure trials. We began by fitting a mixed-effects logistic regression for each age group separately. In all

Condition	Age Group	N	N girls	Mean Age (years)	Age Range
Plausible	3	27	16	3.50	3.00–3.93
Implausible	3	21	8	3.49	3.02–3.93
Plausible	4	21	7	4.60	4.22-4.97
Implausible	4	28	23	4.51	4.00–4.94
Plausible	5	23	19	5.49	5.01–5.95
Implausible	5	20	12	5.48	5.05–5.90

Table 1

Demographic information for participants in Experiment 2.

three age groups, children were more likely to select the plausible referent when asked for it by the Plausible speaker ($\beta_{3-years} = 1.43$, $z = 3.71$, $p < .001$, $d = 1.20$; $\beta_{4-years} = 3.10$, $z = 6.63$, $p < .001$, $d = 2.66$; $\beta_{5-years} = 5.55$, $z = 7.39$, $p < .001$, $d = 8.27$). We then asked whether children's behavior changed over development. A mixed-effects model fit to all of the children's data showed main effects of both age ($\beta_{age} = -.74$, $z = -4.18$, $p < .001$) and Speaker Plausibility ($\beta_{Plausible} = -6.16$, $z = 1.81$, $p < .001$), and also an interaction between the two ($\beta_{age \cdot Plausible} = 2.12$, $z = 7.61$, $p < .001$). Thus, older children showed a greater sensitivity to the speaker on the unambiguous Exposure trials.

We next turned to the Test trials. When we examined each age group separately we found that as in Experiment 1, both 4- and 5-year olds leveraged their previous experience with the speaker when interpreting the ambiguous test utterances (Figure 3). However, 3-year-olds did not ($\beta_{3-years} = .06$, $z = .21$, $p = .83$, $d = .06$; $\beta_{4-years} = 1.01$, $z = 2.87$, $p < .01$, $d = .82$; $\beta_{5-years} = 1.35$, $z = 3.96$, $p < .001$, $d = 1.24$). A mixed-effects regression fit to all of the data confirmed this developmental change in sensitivity, finding a significant main effect of age ($\beta_{age} = -.43$, $z = -2.75$, $p < .01$), a marginal effect of Plausible Speaker ($\beta_{Plausible} = -1.71$, $z = -1.82$, $p = .07$), and a significant interaction between the two

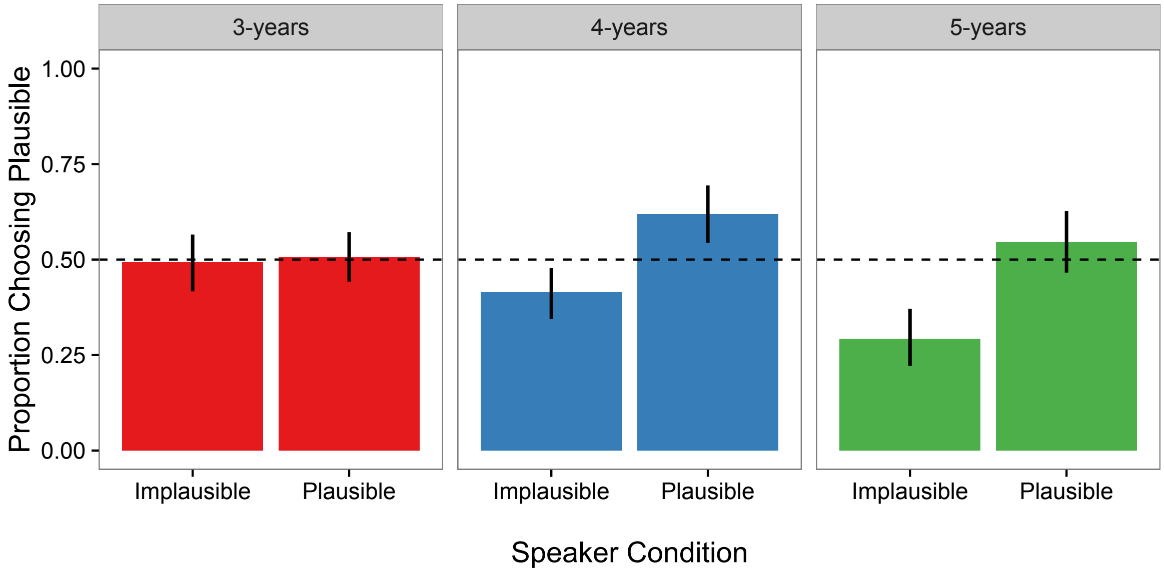

Figure 3. Experiment 2 Test trial responses. Replicating Experiment 1, 4- and 5-year-old children were more likely to correct the phonological error during Test trials when they were previously exposed to a Plausible speaker, although 3-year-olds were not. Bars show group-averaged proportion of plausible item selection at Test, error bars show 95% confidence intervals computed by non-parametric bootstrap at the subject level. The dashed line indicates chance performance.

($\beta_{age \cdot Plausible} = .54$, $z = 2.58$, $p = .01$). In line with previous work, these results show that 3-year-old children have trouble using top-down speaker expectations when processing ambiguous utterances (Kidd & Bavin, 2005). However, children appear to improve significantly over the next two years (Rabagliati, Pylkkänen, & Marcus, 2013).

Older children were thus more sensitive to the speaker's utterances on unambiguous Exposure trials, and relied more on their speaker-expectations on ambiguous Test trials. Did older children rely more on their speaker-expectations because they had built stronger expectations on Exposure trials, or did they they use these expectations differently? To answer this question, we fit an additional model including the proportion of Exposure trials on which individual children had selected the plausible referent. This model showed a significant effect of Plausible Speaker ($\beta_{Plausible} = 1.92$, $z = 3.39$, $p = .001$), a significant effect of Exposure ($\beta_{Exposure} = 2.84$, $z = 4.60$, $p < .001$), and a significant interaction between Plausible Speaker and Exposure ($\beta_{Plausible \cdot Exposure} = -3.32$, $z = -3.64$, $p < .001$), but no effect of age ($\beta_{age} = .06$, $z = .56$, $p = .57$). No other interaction terms improved the model, and this model explained more variance in the data than any other single-interaction models. Thus, it appears that older children relied more on speaker expectations because they had formed stronger expectations rather than because they rely on expectations differently. Experiments 1 and 2 thus show that children's reliance on speaker-expectations remains relatively constant across the 3–6 year range, but that their ability to build these expectations improves gradually across development.

Experiment 3

Experiment 3 tests a second prediction of noisy channel processing: As speech becomes noisier, and thus less reliable, children should rely more on their expectations.

Method

Participants. For Experiment 3, children were again recruited from the floor of the San Jose Children's Discovery museum. As 3-year-olds did not perform differently from chance in Experiment 2, we again focused on 4- and 5-year olds. Data from a total of 114 children were collected, 1 of whom was excluded for < 50% exposure to English and 2 of whom were excluded for parent-reported developmental disabilities. As before, children were recruited until at least 20 had been run in each condition. Children's ages and sexes were again comparable across conditions (Table 2).

Stimuli, Design, & Procedure. Stimuli were constructed as in Experiments 1 and 2 with a few small changes. First, one additional Test trial was added to increase power. Second, two versions of each acoustic recording were made. One was recorded in a sound-proof room by a female native English speaker. The second was constructed by convolving each recording with randomly generated Brown noise with an amplitude of .7

Speaker Condition	Noise Condition	N	N girls	Mean Age (years)	Age Range
Plausible	No Noise	21	8	4.89	4.00–5.83
Plausible	Noisy	26	12	4.89	4.02–5.94
Implausible	No Noise	20	12	4.98	4.00–5.83
Implausible	Noisy	24	11	5.01	4.01–5.92
Implausible (Control)	Noisy	20	11	4.96	4.15–5.91

Table 2

Demographic information for participants in Experiment 3.

using Audacity. The first was used in the No Noise conditions, the second was used to replicate the two conditions from Experiment 1 as well as for the Control condition.

In addition, because all of the Test trials asked the listener to select the semantically implausible referent, it is possible that exposure to an Implausible speaker induced listeners to select the implausible referent at test generally, independent of their acoustic input. In Experiment 3, we provide a Control condition in which the Implausible speaker from Exposure asked for the semantically plausible referent at test. If children inferred from the unambiguous Exposure trials that the goal of the game was to pick the "silly" referent, they should have continued to select the implausible referent on Test trials. In contrast, if Exposure trials caused them to adjust the relative weights on acoustic information and speaker-expectations, they should instead have selected the plausible referent at test.

Results

As in the previous experiments, we first establish that children understood the task, and encoded the differences between the Plausible and Implausible speakers on Exposure trials. At each noise level, children in the Plausible speaker condition were more likely to select the plausible referent on Exposure trials than those in the Implausible speaker condition ($\beta_{NoNoise} = 6.92$, $z = 6.15$, $p < .001$, $d = 7.94$; $\beta_{Noisy} = 4.91$, $z = 8.26$, $p < .001$,

$d = 4.46$). Further, children in the Implausible Control condition performed no differently from those in the Implausible condition, licensing comparison of their Test trials ($\beta_{Control} = .35$, $z = .81$, $p = .42$, $d = .21$).

Second, we again compared the conditions to each other, fitting a mixed-effects regression predicting choice on Exposure trials from noise level and speaker condition (Implausible, Plausible, Control). Compared to the Implausible Speaker condition, children exposed to the Plausible Speaker were more likely to pick the plausible referent on Exposure trials ($\beta_{Plausible} = 6.62$, $z = 8.88$, $p < .001$). Children exposed to the Control speaker did not perform differently on Exposure trials from those exposed to the Implausible speaker, as predicted ($\beta_{Control} = .36$, $z = .83$, $p = .42$). Further, the model showed a marginal effect of Noise level ($\beta_{Noisy} = .86$, $z = 1.74$, $p = .08$) and a significant interaction between Noise level and Speaker ($\beta_{Plausible \cdot Noisy} = -1.71$, $z = -2.15$, $p = .03$), indicating that the addition of noise moved children in both conditions closer to chance.

Did children integrate the noise level of the acoustic stimuli with speaker expectations on the ambiguous Test trials? Figure 4 shows the proportion of trials on which children selected the plausible referent at Test in both speaker and noise conditions. As predicted, children showed sensitivity to both speaker reliability and acoustic noise. Children selected the plausible referent at Test, correcting the error in their acoustic input, more often when the speaker had said plausible things on Exposure trials ($\beta_{Noisy} = 1.17$, $z = 3.72$, $p < .001$, $d = 1.17$; $\beta_{NoNoise} = 1.57$, $z = 4.14$, $p < .001$, $d = 1.17$). In addition, regardless of Speaker plausibility, children selected the plausible referent more frequently when the acoustic input was noisy. To quantify this pattern, we again fit a mixed-effects regression predicting choice on test trials from Speaker type and Noise level as well as their interaction. As predicted, both main effects were significant ($\beta_{Plausible} = 1.96$, $z = 3.20$, $p < .001$, $\beta_{Noisy} = .81$, $z = 2.21$, $p = .03$), but their interaction was not ($\beta_{Plausible \cdot Noisy} = .33$, $z = .65$, $p = .51$).

Finally, one alternative explanation for the difference between Speaker conditions is that children simply followed their expectations at all times, e.g., that those exposed to the

Implausible speaker chose "silly" responses regardless of the question. To test this alternative, we asked whether children who responded to an Implausible speaker on Exposure trials always chose the implausible referent on Test trials even when the speaker referred to the *plausible* referent (Implausible Control condition). A mixed-effects model estimating plausible referent selection on Test trials showed that compared to the Plausible speaker, children exposed to the Implausible speaker were less likely to pick the plausible referent ($\beta_{Implausible} = -1.57$, $z = -4.14$, $p < .001$; $d = 1.17$), but children in the Control condition were *more* likely to select the plausible referent ($\beta_{Control} = 1.06$, $z = 2.05$, $p = .04$; $d = 1.93$). Thus, children who were asked for the plausible referent at Test selected it, even when the speaker had previously always referred to the implausible referent. This Control condition provides further evidence that children were attending to and responding to the acoustic input from the speaker on Test trials, integrating it with their prior expectations.

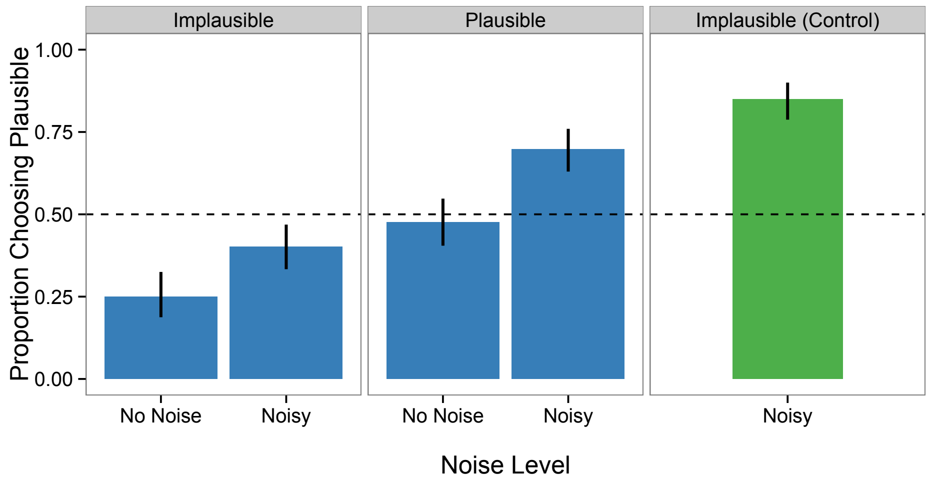

Figure 4. Experiment 3 Test trial responses. Replicating Experiments 1 and 2, children were more likely to correct the phonological error during Test trials when they were previously exposed to a Plausible speaker. In addition, in both conditions children were more likely to correct the error when the speech was noisier. Finally, when children were tested in a Control condition, in which a previously Implausible speaker referred to *plausible* referents during ambiguous Test trials, they reliably selected the plausible referent. Together, these results show that children adaptively integrate both bottom-up acoustic information and top-down expectations about the speaker when processing language. Bars show group-averaged proportion of plausible item selection at test, error bars show 95% confidence intervals computed by non-parametric bootstrap at the subject level. The dashed line indicates chance performance.

General Discussion

When we use language to communicate, we are doing more than processing the sounds we hear; we are trying to infer speakers' intended meaning (H. H. Clark, 1996). Because perception is inherently uncertain, our expectations about what speakers are likely to say play an important role in resolving interpretive ambiguities (Grice, 1975; Frank & Goodman, 2012). Our experiments show that children are able to integrate speaker expectations with perceptual uncertainty by ages 4–5, though perhaps not earlier.

In our studies, children adjusted their reliance on expectations as much as adults, but also generally relied more on top-down expectations. Because of the greater noise inherent in children's perceptual processing systems, the same acoustic stimulus may effectively be less reliable for children than adults (Neuman & Hochberg, 1983; Lyons & Ghetti, 2011). Perhaps children with impaired acoustic processing rely relatively more on their expectations, whereas children with impaired higher-level linguistic expectations rely more on the acoustic signal. Our paradigm could be used to test this prediction.

How much should listeners rely on the acoustic signal and how much should they rely on their expectations? Ideal observer models predict that cue should be weighted in proportion to their reliability (Jacobs, 1999; Ernst & Banks, 2002). This prediction holds for adult listeners across levels in language processing from phonology to syntax (e.g., Gibson et al., 2013; McClelland, Mirman, & Holt, 2006). In our experiments, we cannot say that children's weighting was optimal relative to cue reliability, only that it was adaptive: weighting changed with manipulations of reliability (as in Gibson et al., 2013). It is a challenge for future work to derive independent measures of reliability for high-level linguistic stimuli. Further, because our participants adapted to only one speaker, we cannot know to what extent this adaptation was speaker-specific vs. speaker-general. However, an attractive feature of the noisy channel framework is that it can be applied hierarchically, with appropriate adaptation predicted at the level of speaker, community, and lexicon as evidence accumulates (Kleinschmidt & Jaeger, 2015).

The results of our experiments show that children, like adults, flexibly trade off between information sources in language comprehension in response to their reliability. Noisy channel principles thus provide a framework for understanding language processing in both adults and children.

Authorship

D. Y., S.C., and M.C.F. designed the study. D.Y. and S.C. collected the data. D.Y., S.C., and M.C.F. analyzed the data. D.Y. and M.C.F. wrote the manuscript. All authors approved the final version of the manuscript for submission.

Acknowledgments

We thank Nicolette Castro, NIH NRSA F32HD075577 to DY, and a John Merck Scholars Fellowship to MCF.

References

Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development*, i–174.

Clark, E. V. (2009). *First language acquisition*. Cambridge University Press.

Clark, H. H. (1996). *Using language*. Cambridge University Press.

Clayards, M., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition*, *108*, 804–809.

Creel, S. C. (2012). Preschoolers' use of talker information in on-line comprehension. *Child Development*, *83*, 2042–2056.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*, 429–433.

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*, 998.

Gibson, E., Bergen, L., & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences*, *110*, 8051–8056.

Gori, M., Del Viva, M., Sandini, G., & Burr, D. C. (2008). Young children do not integrate visual and haptic form information. *Current Biology*, *18*, 694–698.

Graham, S. A., Sedivy, J., & Khu, M. (2014). That's not what you said earlier: Preschoolers expect partners to be referentially consistent. *Journal of Child Language*, *41*, 34–50.

Grice, H. P. (1975). Logic and Conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics volume 3: Speech acts* (pp. 41–58).

Jacobs, R. A. (1999). Optimal integration of texture and motion cues to depth. *Vision Research*, *39*, 3621–3629.

Jaeger, F. T. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, *61*, 23–62.

Jelinek, F. (1976). Continuous speech recognition by statistical methods. *Proceedings of the IEEE*, *64*, 532–556.

Kidd, E., & Bavin, E. L. (2005). Lexical and referential cues to sentence interpretation: An investigation of children's interpretations of ambiguous sentences. *Journal of Child Language*, *32*, 855–876.

Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, *122*, 148–203.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*, 1126–1177.

Lew-Williams, C., & Fernald, A. (2007). Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science*, *18*, 193–198.

Lyons, K. E., & Ghetti, S. (2011). The development of uncertainty monitoring in early childhood. *Child Development*, *82*, 1778–1787.

Matthews, D., Lieven, E., & Tomasello, M. (2010). What's in a manner of speaking? Children's sensitivity to partner-specific referential precedents. *Developmental Psychology*, *46*, 749–760.

McClelland, J. L., Mirman, D., & Holt, L. L. (2006). Are there interactive processes in speech perception? *Trends in Cognitive Sciences*, *10*, 363–369.

Nardini, M., Bedford, R., & Mareschal, D. (2010). Fusion of visual cues is not mandatory in children. *Proceedings of the National Academy of Sciences*, *107*, 17041–17046.

Nardini, M., Jones, P., Bedford, R., & Braddick, O. (2008). Development of cue integration in human navigation. *Current Biology*, *18*, 689–693.

Neuman, A. C., & Hochberg, I. (1983). Children's perception of speech in reverberation. *The Journal of the Acoustical Society of America*, *73*, 2145–2149.

Pasquini, E. S., Corriveau, K. H., Koenig, M., & Harris, P. L. (2007). Preschoolers monitor the relative accuracy of informants. *Developmental Psychology*, *43*, 1216–1226.

Rabagliati, H., Pylkkänen, L., & Marcus, G. F. (2013). Top-down influence in young children's linguistic ambiguity resolution. *Developmental Psychology*, *49*, 1076–1089.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell Systems Technical Journal*, *XXVII*, 379–423.

Snedeker, J., & Trueswell, J. C. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology*, *49*, 238–299.

Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition*, *73*, 89–134.