

Preschoolers flexibly adapt to noisy linguistic input

Daniel Yurovsky, Sarah Case, and Michael C. Frank

Stanford University

Author Note

Please address correspondence to:

Daniel Yurovsky

Jordan Hall (Building 420)

Stanford University

450 Serra Mall

Stanford, CA 94305

Email: yurovsky@stanford.edu

Word Count:

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 4- and 5-year-old children with ambiguous utterances that afforded two different interpretations, depending on whether they privileged perceptual input or linguistic expectations. Across two 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, children flexibly adjusted their interpretations as cues changed in reliability.

Keywords: Language processing, noisy channel, cognitive development

Preschoolers flexibly adapt to noisy linguistic input

Imagine Bob hears Alice say “I had carrots and bees for dinner.” Perhaps she visited an exotic restaurant, and he should follow up and ask how they tasted. Or perhaps he misheard her or she misspoke, and she actually ate *peas*. To interpret Alice’s utterance, Bob has to integrate perceptual information from her speech with his expectations about both what words usually go with “carrots” and “dinner” and also what foods people usually eat. Modern statistical language processing systems operate using 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 the human language processing system operates in complex and uncertain real-time communicative situations (Clayards, Tanenhaus, Aslin, & Jacobs, 2008; Levy, 2008; Jaeger, 2010; Kleinschmidt & Jaeger, 2015). On this view of language, comprehenders integrate prior expectations with perceptual data; and they do so probabilistically, weighting each according to its reliability (Ernst & Banks, 2002; Jacobs, 1999). In one demonstration of this kind of 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 (e.g., the omission of the word “to”). Adult comprehenders were able to correct these errors, and critically corrected them at higher rates when they had reason to believe that the communicative channel was noisy (and hence the perceptual signal was unreliable). Conversely, they corrected these errors at lower rates when they believed they were in a “silly” context where many of the sentences they read were similarly implausible. Together, these results suggested a flexible weighting of information from the perceptual signal and from speaker expectations.

Do children also process language in the same flexible, expectation-based way? Such

flexibility would be important for making efficient use of input in complex and varied linguistic environments. In addition, developmental evidence on noisy-channel processing contributes to an understanding of the origins of processing ability. Though to our knowledge no experiments directly test noisy channel models in childhood, relevant prior work is mixed.

On one hand, infants and toddlers already make use of speakers' social and pragmatic cues to reference to shape their expectations about linguistic meaning (Carpenter, Nagell, Tomasello, Butterworth, & Moore, 1998; E. V. Clark, 2009; Creel, 2012). On the other hand, children sometimes show striking deficits in revising incorrect expectations about meaning in cases where syntax is ambiguous (Trueswell, Sekerina, Hill, & Logrip, 1999). These failures are seen, however, in complex contexts that were designed to test the development of other abilities. Our current experiments attempt a test of expectation-based integration in the absence of other processing demands.

We created a paradigm that allowed independent manipulation of both expectations about a speaker's plausibility and perceptual noise. We introduced four- and five-year-old children (and, as a control group, adults) to either a plausible or implausible speaker, who initially uttered unambiguous sentences like "my cat has three little [kittens/hammers]" (Figure 1a). We then asked participants to resolve the intended meaning for ambiguous sentences such as the "bees/peas" example above, which could either be the product of a mishearing or phonological error, or could convey implausible content (Figure 1b). We predicted that, if children are able to integrate speaker expectations and channel noise, then their interpretations should be a product of both of these. In Experiment 1, we provide a first test of this hypothesis by manipulating speaker plausibility. In Experiment 2, we replicate the findings of Experiment 1 and cross these with a manipulation of channel noise. The results of these two experiments show that children, like adults, flexibly trade off between information sources in language comprehension in response to their reliability.

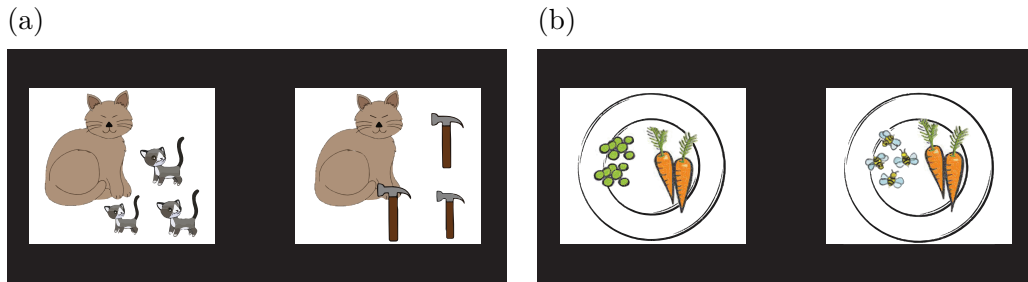


Figure 1. Example Exposure and Test trials. On Exposure trials, the two pictures and their corresponding referring expressions were highly distinct (e.g. “my cat has three little [kittens/hammers].” In contrast, the two pictures on Test trials could be referred to by expressions that contained a single phonological error (e.g. I had carrots and [bees/peas] for dinner.” For participants in the Implausible condition, the speaker referred to the implausible referent (e.g. “hammers”) on all Exposure trials. For participants in the Plausible condition, the speaker referred to to the plausible referent on all Exposure trials (e.g. “kittens”). On test trials, the speaker always referred to the implausible referent (“bees”).

Experiment 1

Method

Participants. Adult participants for Experiment 1 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.

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. 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)).

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.¹ 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 a small amount of Brown noise.

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 eight 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 this introduction, 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.²

¹All stimuli, as well as data and analyses are available in a public repository at <http://github.com/dyurovsky/noisy-kids>.

²For a small subset of participants (N=3), a software bug meant that responses were not tracked auto-

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 chosen the plausible referent (e.g. kittens) and those in the Implausible condition should have chosen the implausible referent (e.g. hammers). We provide two convergent analysis that show that they did.

First, we asked whether participants selected the correct referent in each condition at levels different from chance. We submitted performance for each age group and condition separately to a mixed-effects logistic regression with random effects of subject and trial. Both children and adults were significantly more likely to choose the plausible referent in the Plausible condition ($\beta_{adult} = 5.66, z = 6.76, p < .001$; $\beta_{child} = 2.62, z = 5.43, p < .001$), and significantly less likely than chance to choose the plausible referent in the Implausible condition ($\beta_{adult} = -3.87, z = -7.13, p < .001$; $\beta_{child} = -1.9, z = 5.61, p < .001$). Second, we compared performance across conditions, fitting a mixed-effects regression predicting choice on Exposure trials from group, condition, and their interaction, and found significant effects of all three ($\beta_{child} = 1.96, z = 3.33, p < .001$, $\beta_{PlausibleSpeaker} = 9.52, z = 7.87, p < .001$, $\beta_{child \times condition} = -5, z = -4.12, p < .001$). 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. While children were overall more likely to pick the plausible referent in both conditions, the effect size of the difference between conditions

matically and were post-coded from videotapes.

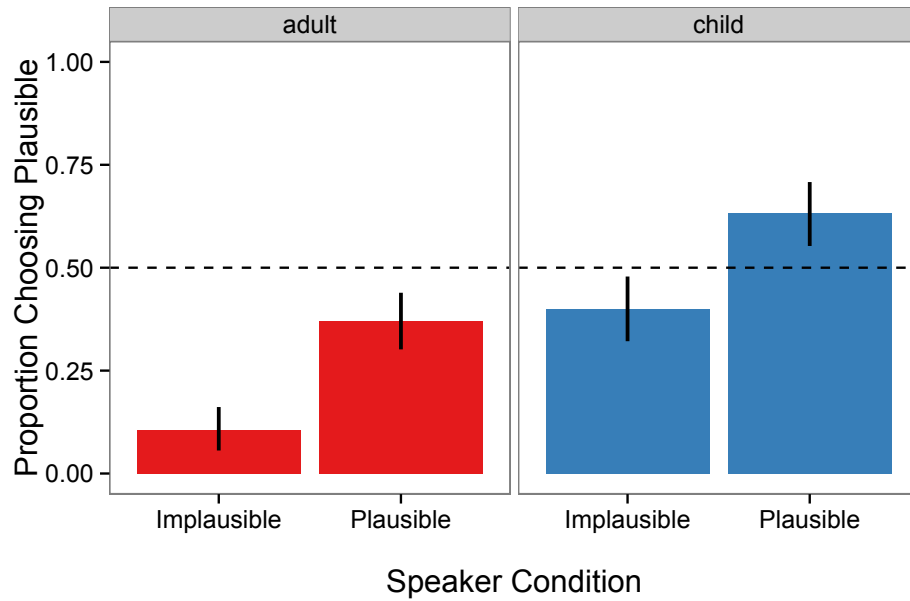


Figure 2. Experiment 1 test trial selections for both children and adults. In line with the predictions of a noisy channel model of language processing, both children and adults were more likely to correct the phonological error during test trials when they were previously exposed to a plausible speaker.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

was nearly identical in adults and children, indicating equal adaptation to the Plausible vs. Implausible speaker ($d_{adult} = 1.09$, $d_{child} = 1.10$). 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, indicating that both children and adults 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 ($\beta_{child} = 1.96$, $z = 3.33$, $p < .001$, $\beta_{PlausibleSpeaker} = 2.33$, $z = 4.40$, $p < .001$, $\beta_{child \times condition} = -1.01$, $z = -1.53$, $p = .13$).

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 preschoolers 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 1 shows that children, to the same degree as adults, flexibly adapt their on-line language processing in accord with their expectations about what a speaker is likely to say. In Experiment 2, we replicate this finding and also test a further prediction of the noisy channel framework for language processing: children should adapt to the reliability of the acoustic signal. As speech becomes noisier, and thus harder to hear, listeners should rely more on their expectations.

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 2, we provide a Control condition in which the Implausible speaker from training 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 linguistic expectations, they should instead have selected the plausible referent at test.

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 114 children were collected, 1 of whom was excluded for insufficient exposure to English and 2 who 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 (Plausible, Noisy – 26 children, 12 girls, age: mean = 4.98, min = 4.02, max = 5.94; Implausible, Noisy – 24 children, 11 girls, age: mean = 5.01, max = 5.92, min = 4.01, Plausible, No Noise = 21 children, 8 girls, age: mean = 4.89, max = 5.93, min = 4.1, Implausible, No Noise — 20 children, 12 girls, age: mean = 4.98, max = 5.83, min = 4, Control — 20 children, 11 girls, age: mean = 4.96, max = 5.91, min = 4.15).

Stimuli, Design, & Procedure. Stimuli were constructed as in Experiment 1 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 as before. The second was constructed by convolving each recording with randomly generated Brown noise with an amplitude of .7. 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.

Results

As in Experiment 1, we first establish that children understood the task, and encoded the differences between the Plausible and Implausible speakers on Exposure trials. First, we ask whether children selected the plausible referent when it was asked for by the Plausible speaker, and selected the implausible referent when it was asked for by the Implausible speaker independent of noise level. We submitted performance for each

condition and noise level separately to a mixed-effects logistic regression with random effects of subject and trial. As before, children were more likely than chance to choose the plausible referent in the Plausible condition, but less likely than chance to choose the plausible referent when the speaker asked for the implausible referent (smallest $\beta_{Control} = -.153$, $z = -4.32$, $p < .001$).

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, who was identical to the Implausible speaker on Exposure trials, were not significantly different, 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 ($\beta_{Plausible \times Noisy} = -1.71$, $z = -2.15$, $p = .03$), 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 3 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 (as in Experiment 1). 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 ($\beta_{1.20} = 1.96$, $z = 3.20$, $p < .001$, $\beta_{Noisy} = .81$, $z = 2.21$, $p = .03$), but their interaction was not ($\beta_{Plausible \times Noisy} = .33$, $z = .65$, $p = .51$).

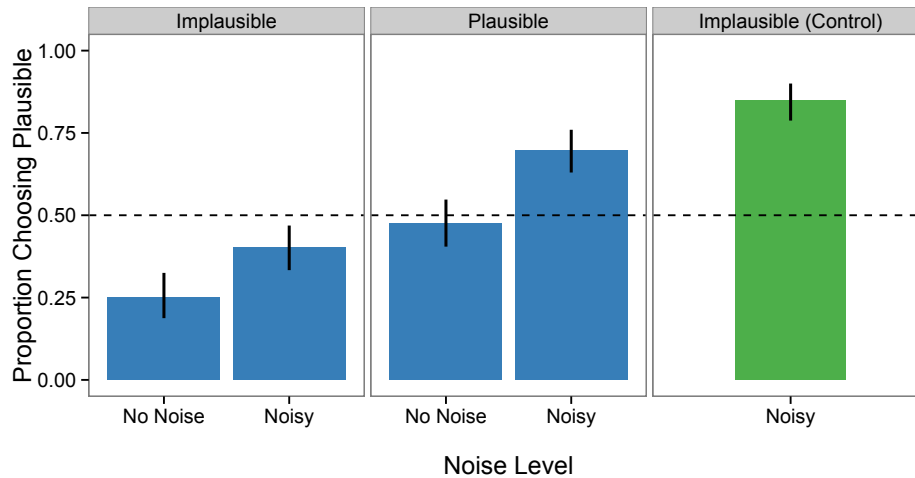


Figure 3. Experiment 2 Test trial responses. Replicating Experiment 1, children were more likely to correct the phonological error during test trials when they were previously exposed to a Plausible speaker. In addition, as predicted, in both conditions children were more likely to correct the error when the speech was noisier. Finally, when children were tested in a control, 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. The dotted line indicates chance performance.

Finally, one alternative explanation for the difference between Speaker conditions is that children simply followed their expectations at all times, e.g., that they 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 ($\beta_{\text{Implausible}} = -1.57$, $z = -4.14$,

$p < .001$), but children in the Control condition were *more* likely to select the plausible referent ($\beta_{Control} = 1.06$, $z = 2.05$, $p = .04$). 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, and not just relying on their prior expectations.

General Discussion

When we use language to communicate, we are doing more than processing the sounds we hear; we are trying to infer the speaker’s intended meaning (H. H. Clark, 1996). Because of the noise and ambiguity inherent in linguistic communication, our expectations about what speakers are likely to say play an important role in resolving interpretive ambiguities (Grice, 1975; Frank & Goodman, 2012). But how much should listeners rely on the acoustic signal and how much should they rely on their expectations? Ideal observer models of this kind of cue integration predict that each cue should be weighted in proportion to its reliability (Jacobs, 1999; Ernst & Banks, 2002). This prediction is borne out in adult listeners across domains in language processing from phonology to syntax (e.g., Gibson et al., 2013; McClelland, Mirman, & Holt, 2006; Kleinschmidt & Jaeger, 2015). The experiments in this paper provide evidence that this adaptive expectation-based language processing is already available to 4–5 year old children. Not only do preschoolers combine their expectations about a speaker’s likely utterance with the speech they are hearing, they flexibly re-weigh their reliance on each as they develop stronger expectations about the speaker and as the speech signal changes in clarity.

In our data, while the size of the adaptation effect was nearly identical in both children and adults, their baseline reliance on perceptual vs. linguistic cues was not. Children relied more on their top-down expectations than the perceptual signal. Because of the greater noise inherent in young children’s perceptual processing systems, the same

acoustic stimulus may effectively be less reliable for children than adults (Neuman & Hochberg, 1983; Lyons & Ghetti, 2011). Relatedly, the noisy channel framework makes a clear prediction about integration in atypical populations: Children whose acoustic processing is impaired (e.g., with cochlear implants) should rely relatively more on their expectations, whereas children with impaired higher-level linguistic expectations (e.g., children with specific language impairment) should rely more the acoustic signal.

Our work here suggests that the noisy channel framework offers a route to understanding both children’s and adults’ language processing, as well as how one develops from the other. Within this framework, processing is characterized by the interaction between two models: one of perception, and one of expectation. The content and level of description for these models are outstanding questions, however. One possibility is that expectations are fundamentally about speakers, and that linguistic communication is essentially pragmatic all the way down. Alternatively, these models could be linguistic in content, about the distributions of words, referents, and their mutual predictability. We hope that the paradigm we have developed here will be a tool for helping to resolve this fundamental tension.

Acknowledgments

Thanks to Nicolette Castro for data collection. This work was supported by 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, Sep). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition*, 108(3), 804-9.
- Creel, S. C. (2012). Preschoolers' use of talker information in on-line comprehension. *Child Development*, 83(6), 2042–2056.
- Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(6870), 429–433.
- Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, 336(6084), 998.
- Gibson, E., Bergen, L., & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. In *Proceedings of the . . .*
- 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*.
- Jaeger, F. T. (2010, August). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, 61(1), 23–62.
- Jelinek, F. (1976). Continuous speech recognition by statistical methods. In *Proceedings of the ieee* (pp. 532–556).
- Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*.
- Levy, R. (2008, March). Expectation-based syntactic comprehension. *Cognition*, 106(3),

1126–1177.

Lyons, K. E., & Ghetti, S. (2011). The development of uncertainty monitoring in early childhood. *Child development*, 82(6), 1778–1787.

McClelland, J. L., Mirman, D., & Holt, L. L. (2006, August). Are there interactive processes in speech perception? *Trends in Cognitive Sciences*, 10(8), 363–369.

Neuman, A. C., & Hochberg, I. (1983). Children’s perception of speech in reverberation. *The Journal of the Acoustical Society of*

Shannon, C. E. (1948). A mathematical theory of cummunication. *The Bell Systems Technical Journal*, XXVII(3), 379–423.

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(2), 89–134.