



ELSEVIER

Contents lists available at ScienceDirect

## Journal of Memory and Language

journal homepage: [www.elsevier.com/locate/jml](http://www.elsevier.com/locate/jml)

## Talker-specificity and adaptation in quantifier interpretation

Ilker Yildirim<sup>a,b,c,\*</sup>, Judith Degen<sup>d</sup>, Michael K. Tanenhaus<sup>a,e</sup>, T. Florian Jaeger<sup>a,e,f,\*</sup><sup>a</sup> University of Rochester, Department of Brain and Cognitive Sciences, United States<sup>b</sup> Massachusetts Institute of Technology, Department of Brain and Cognitive Sciences, United States<sup>c</sup> The Rockefeller University, Laboratory of Neural Systems, United States<sup>d</sup> Stanford University, Department of Psychology, United States<sup>e</sup> University of Rochester, Department of Linguistics, United States<sup>f</sup> University of Rochester, Department of Computer Science, United States

## ARTICLE INFO

## Article history:

Received 6 August 2014

revision received 7 August 2015

Available online xxxx

## Keywords:

Adaptation

Talker-specificity

Quantifiers

Semantics

Pragmatics

## ABSTRACT

Linguistic meaning has long been recognized to be highly context-dependent. Quantifiers like *many* and *some* provide a particularly clear example of context-dependence. For example, the interpretation of quantifiers requires listeners to determine the relevant domain and scale. We focus on another type of context-dependence that quantifiers share with other lexical items: talker variability. Different talkers might use quantifiers with different interpretations in mind. We used a web-based crowdsourcing paradigm to study participants' expectations about the use of *many* and *some* based on recent exposure. We first established that the mapping of *some* and *many* onto quantities (candies in a bowl) is variable both within and between participants. We then examined whether and how listeners' expectations about quantifier use adapts with exposure to talkers who use quantifiers in different ways. The results demonstrate that listeners can adapt to talker-specific biases in both how often and with what intended meaning *many* and *some* are used.

© 2015 Elsevier Inc. All rights reserved.

## Introduction

The meaning of *many*, if not all, words is context-dependent. For example, whether we want to say that John is *tall* depends on whether John is being compared to other boys his age, professional basketball players, dwarves, etc. (e.g., Halff, Ortony, & Anderson, 1976; Kamp, 1995; Kennedy & McNally, 2005; Klein, 1980). Other words whose interpretation requires reference to context are pronouns and quantifiers (Bach, 2012). For example, the interpretation of a quantifier like *many* depends on the class of objects that is being quantified over: the number of

crumbs that *many crumbs* refers to is judged to be higher than the number of mountains that *many mountains* refers to (Hörmann, 1983).

A less-well studied aspect of context-dependence is how a given talker uses quantifiers like *many* and *some*. Talkers exhibit individual variability at just about any linguistic level investigated – including, for example, pronunciation (e.g., Allen, Miller, & DeSteno, 2003; Bauer, 1985; Harrington, Palethorpe, & Watson, 2000; Jaeger-Dror, 1994), lexical preferences (e.g., Finegan & Biber, 2001; Roland, Dick, & Elman, 2007; Tagliamonte & Smith, 2005), and syntactic preferences (e.g., the frequency with which they use passives, Weiner & Labov, 1983). Therefore, talkers are also likely to differ in how they use quantifiers. For example, talkers may differ in how many crumbs they consider to be *many crumbs*, and these differences would consequently be reflected in their productions. In this case, listeners would be well served by taking into account

\* Corresponding authors at: University of Rochester, Department of Brain and Cognitive Sciences, United States.

E-mail addresses: [ilkery@mit.edu](mailto:ilkery@mit.edu) (I. Yildirim), [fjaeger@bcs.rochester.edu](mailto:fjaeger@bcs.rochester.edu) (T.F. Jaeger).

<http://dx.doi.org/10.1016/j.jml.2015.08.003>

0749-596X/© 2015 Elsevier Inc. All rights reserved.

talker-specific knowledge in order to successfully infer what the talker intended to convey.

Talker-specific knowledge has been observed experimentally in cases of variation in pronunciation and syntactic production (e.g., Clayards, Tanenhaus, Aslin, & Jacobs, 2008; Creel & Bregman, 2011; Creel & Tumlin, 2009; Fine, Jaeger, Farmer, & Qian, 2013; Kamide, 2012; Kraljic & Samuel, 2007). While this question has received less attention in lexical processing, there is some evidence that listeners can learn to anticipate talker-specific biases in the frequency with which referents are being referred to (Metzing & Brennan, 2003) and that these talker-specific expectations are reflected in online processing (e.g., Creel, Aslin, & Tanenhaus, 2008). These studies complement classic work on *conceptual pacts* in which interlocutors adjust their use of referential expressions to create temporary, shared context-specific names (Brennan & Clark, 1996).

Previous work on talker-specific lexical expectations has focused on open class, semantically rich, content words – typically nouns (Brennan & Clark, 1996; Creel et al., 2008; Metzing & Brennan, 2003). This raises the question of whether listeners are capable of adapting to talker-specific differences in the use of words that convey more abstract meanings, such as those of quantifiers. If listeners do in fact adapt to talker-specific differences, what specifically are listeners adapting to, i.e., what is the nature of the representations that are being updated and what are the underlying mechanisms?

The current paper begins to address these questions by studying adaptation to talker-specific differences in the use of the quantifiers *some* and *many*. We present four experiments that investigate lexical adaptation. Taken together, these experiments establish (i) that listeners *can* adapt to talker-specific differences in the usage of even abstract lexical items, such as quantifiers; (ii) that, provided sufficient exposure, such adaptation can be achieved even for multiple talkers simultaneously; (iii) that lexical adaptation is observed both to talker-specific differences in the frequency with which lexical items are used *and* to talker-specific differences in how they are being used; and thus, finally, (iv) that lexical adaptation – although often studied as a separate phenomena—exhibits many of the hallmarks of adaptation observed for other linguistic domains. Next, we elaborate on these points, while introducing the four experiments presented below. In doing so, we relate our research to previous work and highlight the contributions of the current work.

Before we investigate lexical adaptation to talker-specific quantifier use, we first assess whether the premise for adaptation is given: Experiment 1 demonstrates that listeners differ in their initial expectations about a talker's use of a variety of quantifiers, including *some* and *many*. This shows that if listeners want to arrive at an interpretation of an utterance that is close to the talker's intended meaning, they might sometimes need to adapt their expectations about quantifier use to match those of the current talker. Experiment 1 thus provides the first direct evidence that there would potentially be a *benefit* to adaptation to talker-specific differences in quantifier use.

This then raises the question whether listeners do adapt to these changes. This is the central motivation for

Experiment 2. Going beyond this question and previous work, Experiment 2 also begins to investigate the nature of the changes in expectations that result from exposure to a novel talker. Specifically, we ask whether lexical adaptation can be talker-specific. The answer to this question is of theoretical relevance, as it speaks to the nature of the mechanisms underlying lexical adaptation. We briefly elaborate on this point, as it has so far received relatively little attention in the literature on lexical adaptation (but see Brennan & Clark, 1996; Pickering & Garrod, 2004).

A priori, there are several ways in which a listener can treat experience with a novel talker. A listener might treat new experience as evidence that can be used to sharpen prior expectations about quantifier use without taking into account the specific context, including the talker. Any adaptation would then be to talkers in general. At the other extreme, adaptation might be completely context-specific. If that were the case, then adaptation would be specific to a particular talker in a particular context and would not at all generalize to other talkers. A more likely possibility is that listeners strike a subtle balance between context-general and context-specific adaptation (cf. Kleinschmidt & Jaeger, 2015). Prima facie, it would seem undesirable for a language processing system to allow a small amount of recent exposure to overwrite life-long experience with language. At the same time, it is beneficial to be able to rapidly adapt to talker-specific lexical preferences, potentially increasing the efficiency of communication (for related discussion, see Brennan & Clark, 1996; McCloskey & Cohen, 1989; McRae & Hetherington, 1993; Pickering & Garrod, 2004; Seidenberg, 1994).

One way to meet both the need for adaptation and the need to maintain previously acquired knowledge is to learn and maintain talker-specific expectations, so that adaptation to a novel talker does not imply loss of previously acquired knowledge. Research in speech perception has explored and found support for this hypothesis (Goldinger, 1996; Johnson, 2006; Kraljic & Samuel, 2007; for review, see Kleinschmidt & Jaeger, 2015). More recent research has found support for this idea in other domains of language processing (e.g., prosodic processing, Kurumada, Brown, Bibyk, Pontillo, & Tanenhaus, 2014; Kurumada, Brown, & Tanenhaus, 2012; and sentence processing, Fine et al., 2013; Jaeger & Snider, 2013). For example, in episodic and exemplar-based models, linguistic experiences are assumed to be stored along with knowledge about the context in which they occurred (Goldinger, 1996; Johnson, 2006; Pierrehumbert, 2001). This is how these models capture talker-specific expectations. (Similar reasoning applies to Bayesian models of adaptation that assume generative processes over hierarchically organized indexical alignment, Kleinschmidt & Jaeger, 2015). Similarly, memory-based models of lexical alignment (Horton & Gerrig, 2005, *in press*) can in theory account for both talker-specific expectations – if talkers are included as contexts (Brown-Schmidt, Yoon, & Ryskin, 2015).

Changes in the use of lexical forms and structures due to exposure are often attributed to temporary changes in expectation within a spreading-activation framework. These “priming-based” accounts assume that exposure

increases the activation of a particular word, structure and perhaps conceptually related words and structures (e.g., Arai, van Gompel, & Scheepers, 2007; Branigan, Pickering, & McLean, 2005; Chang, Dell, & Bock, 2006; Goudbeek & Krahmer, 2012; Pickering & Branigan, 1998; Reitter, Keller, & Moore, 2011; Traxler & Tooley, 2008). Although lexical priming accounts have not been applied to the issues we are exploring, the simplest version of these models would most naturally predict that changes in expectations would apply across talkers and thus be talker-independent. In contrast the models discussed above – while compatible with generalization across talkers—predict there to be also talker-specific expectations (as we discuss later, *some* generalization is, in fact, expected under these alternative accounts).

In Experiment 2a and Experiment 2b, respectively, we ask listeners to either make judgments about “a talker”, which leaves ambiguous the possibility that we are referring to any talker, or “the talker”, referring to the specific talker to whom they were exposed. Comparing the two experiments allow us to ask whether listeners adapt, at least in part, to a specific talker, rather than changing their expectations across the board to reflect how any new talker might use *some* and *many*.

Building on the basic effect observed in Experiments 2a and 2b, Experiment 3 then asks whether listeners adjust not only to changes in the frequency with which quantifiers are used by a given talker, but also to changes in how quantifiers are used to refer to specific quantities by a given talker. Both of these quantities are of theoretical interest: talkers might differ in either or both of these aspects, so that the ability to adapt to such differences is potentially beneficial for listeners. Additionally, if lexical adaptation at least qualitatively follows the principles of rational inference and learning (as has been proposed for phonetic adaptation, Kleinschmidt & Jaeger, 2011, 2015, in press and syntactic adaptation, Fine, Qian, Jaeger, & Jacobs, 2010; Kleinschmidt, Fine, & Jaeger, 2012), listeners are expected to be sensitive to both prior probability of quantifiers (i.e., their frequency of use) and the likelihood of quantifiers given an intended interpretation (i.e., how quantifiers are used). Although not framed in these terms, previous work has exclusively focused on adaptation to changes in the frequency (and only for content words, e.g., Creel et al., 2008; Metzger & Brennan, 2003), leaving open whether listeners can adapt to changes in the likelihood. Experiment 3 tests whether listeners can also adapt to changes in the likelihood.

Finally, in Experiment 4 we return to the question of talker-specificity and ask whether listeners can adapt to multiple talkers simultaneously, when these talkers differ in how they use *some* and *many*. This prediction is made by episodic (Goldinger, 1996), exemplar-based (Johnson, 1997, 2006; Pierrehumbert, 2001), and certain Bayesian models (Kleinschmidt & Jaeger, 2015) of adaptation in speech perception. Talker-specific adaptation to multiple talkers has been observed in experiments on speech perception (Kraljic & Samuel, 2007) and, more recently, during syntactic processing (Kamide, 2012). To the best of our

knowledge, it has not previously been tested for lexical processing. Experiment 4 exposes listeners to two talkers with different usage of *some* and *many*.

The studies presented here thus extends previous research on lexical adaptation and alignment in comprehension both methodologically – by establishing the exposure-test paradigm frequently used in research on speech perception as suitable for research on lexical adaptation—and empirically. We find that listeners can adapt to both how often and with what intended interpretation specific talkers use *some* and *many*, and that – at least in simple situations like those investigated here—listeners can adapt to talker-specific quantifier use of multiple talkers from very little input. The experiments presented here establish a novel paradigm to investigate lexical adaptation in ways parallel to research on adaptation to talker variability in speech perception. This makes our results comparable to research in these other fields. Indeed, we find several parallels between lexical adaptation and adaptation at other levels of language processing. We close by discussing venues for future research on lexical adaptation that, we think, are facilitated by the current paradigm.

### Experiment 1: Variability in quantifier interpretation

It is well-known that there are gradient context-dependent differences in the interpretation of quantifiers (e.g., Hörmann, 1983; Newstead, 1988; Pepper & Prytclak, 1974). It is less clear, however, whether talkers differ in their use of quantifiers. For example, talkers could differ in the overall frequency with which they use a certain quantifier, in their interpretation of a quantifier (i.e., when they will use it), or both. If there is such variation, different listeners – who have been exposed to different talkers—are expected to vary in their assumptions about how quantifiers are used. If there is no such variation, it seems unlikely that there is talker variability and thus there is no reason to expect that listeners should adapt to talker-specific usage of quantifiers. Thus, Experiment 1 seeks to establish whether listeners have different expectations about talkers' usage of quantifiers. As our plan going into Experiment 1 was to investigate talker-specific adaptation in quantifier use in subsequent experiments, we explored listener-specific expectations for five quantifiers, *few*, *many*, *most*, *several*, and *some*.

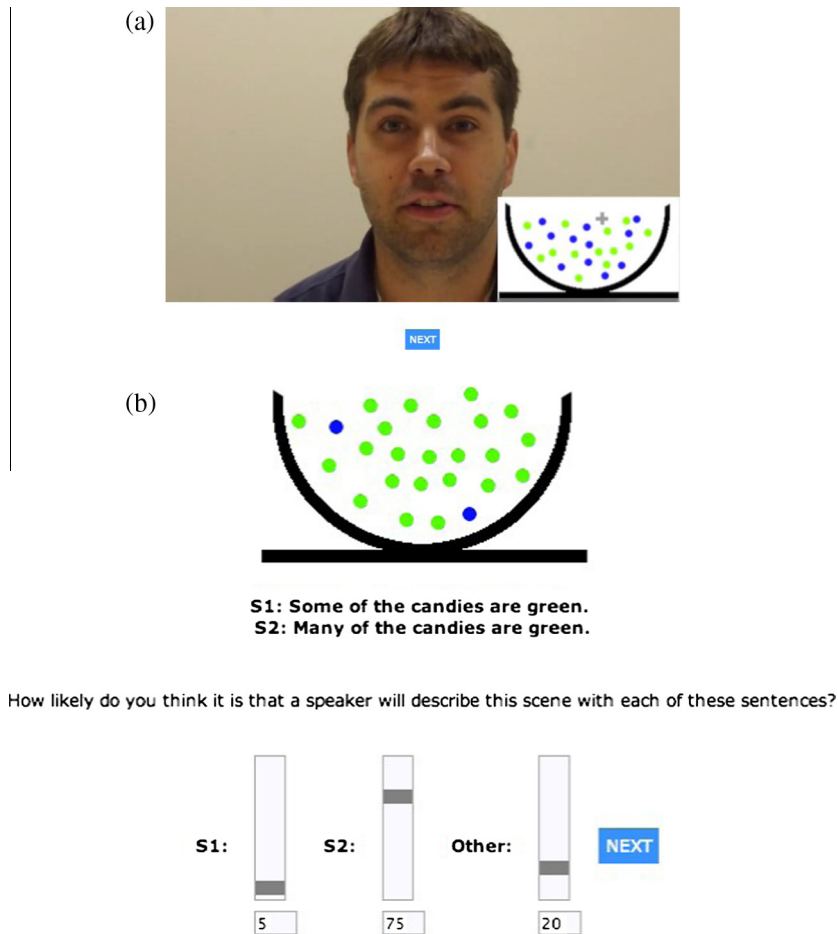
#### Methods

##### Participants

A total of 200 participants were recruited via Amazon's crowdsourcing platform Mechanical Turk (20 per list; see below). All participants were self-reported native speakers of English. The experiment took about 10 min to complete. Participants were paid \$1.00 (\$6.00/h).

##### Materials and procedure

On each trial, participants saw a candy scene in the center of the display (example trial Fig. 1(a)). The bowl always contained a mixture of green and blue candies. The total number of candies in the bowl was constant at 25 but



**Fig. 1.** Panel (a) illustrates the procedure of Experiment 1. The two phases of Experiment 2 are (a) exposure and (b) post-exposure.

the distribution of green and blue candies and the spatial configuration of the candies differed between scenes. At the bottom of the scene, participants saw three alternative descriptions. One of the alternatives was always “Other”. The two other alternatives were two sentences that differed only in their choice of quantifier (e.g., *Some of the candies are green* and *Many of the candies are green*). The alternatives a given participant saw remained the same throughout the experiment. For the five English quantifiers we were interested in (*few*, *many*, *most*, *several*, and *some*), there were ten possible pairwise combinations: (1) *many* and *most*; (2) *many* and *several*; (3) *many* and *few*; (4) *many* and *some*; (5) *most* and *several*; (6) *most* and *few*; (7) *most* and *some*; (8) *several* and *few*; (9) *several* and *some*; and (10) *few* and *some*.

Each participant saw only one of these 10 possible combinations and each combination was seen by equally many participants (20 each). Between participants and within quantifier combinations, the order of presentation of the quantifiers was balanced (e.g., 10 participants saw *Some of the candies are green* on the top and *Many of the candies are green* on the bottom and 10 other participants saw these two sentences in the opposite order).

Participants were asked to rate how likely they thought a talker would be to describe the scene using each of

the alternative descriptions. They performed this task by distributing a total of 100 points across the two alternatives (the first and the second slider bars in Fig. 1(b)) and “Other” to reflect how likely they thought that neither alternative was likely to be used to describe the scene (the third slider bar). Sliders adjusted automatically to guarantee that a total of 100 points were used. An example display for the two quantifiers *some* and *many* is shown in Fig. 1(b).

To assess participants’ beliefs about talkers’ use of all the five quantifiers, we sampled scenes representing the entire scale – a scene could contain any number of green candies from none to 25. Over 78 test trials, participants rated each possible number of green candies 3 times. The order of the scenes was pseudo-randomized, and the mapping from alternative descriptions to slider bars was counterbalanced.

#### Exclusions

To ensure that participants were attending to the task, the experiment contained catch trials after about every 6 trials, totaling 13 catch trials. Catch trial occurrence was randomized so as to rule out strategic allocation of attention. On about half of the catch trials, a gray cross appeared at a random location in the scene. After the scene was



removed from the screen and before the next scene was shown, participants were asked if they had seen a gray cross in the previous scene. In all experiments reported in this paper, we excluded participants who did not respond correctly on at least 75% of the catch trials. We also excluded participants who did not adjust the slider bars for the entirety of the experiment. We excluded five participants out of 200 participants, all on the basis of their catch trial performance: one participant in *some* vs. *many*, one participant in *few* vs. *many*, one participant in *few* vs. *some*, one participant in *many* vs. *several*, and one participant in *several* vs. *some*.

### Results and discussion

In Fig. 2 we show participants' marginal expectations about quantifier use for the five quantifiers. These expectations were obtained by pooling the ratings for each quantifier (e.g., ratings for *some* across the four pairs it appeared), thereby averaging across contrasts (quantifier pairs) and a total of about 80 participants per quantifier.

Analyses revealed considerable individual variation in participants' expectations about the use of these five quantifiers. Here we focus on the assessment of individual variability in participants' expectations about *many* and *some*, the two quantifiers that the rest of the paper will be concerned with. We chose to focus on these two quantifiers, because the paradigm we introduce in Experiment 2 aims to 'shift' listeners' expectations about quantifier use through exposure. We thus focused on quantifiers that – across participants – had peaks in their distributions that were clearly distinct from the edges of our scale (i.e., 1 and 25 candies). Among the three quantifiers that fulfilled this criterion (*some*, *several*, and *many*), we chose to focus on two more frequent ones (*some* and *many*).

We illustrate the variability in listeners' expectations about the use of *some* and *many* by fitting a linear mixed model (Baayen, Davidson, & Bates, 2008) using the *lme4* package (Bates, Maechler, Bolker, & Walker, 2014) in *R* to the data of the 19 participants that rated *many* compared to *some* (recall that one participant was excluded because of poor performance). The distributions of ratings of *some* and *many* (cf. Fig. 2) were separately fit using natural splines (Harrell, 2014) with two degrees of freedom (locations of knots automatically determined using the package *rms*, Harrell, 2014). Random by-participant slopes were included for both of the spline parameters and for the intercepts. The results of this procedure are shown for three representative participants in Fig. 3. This was also evidenced by the estimated variance in the by-participant slopes or the two parameters of the natural splines (e.g., in the case of *many* distributions:  $\sigma_1 = 24.4$ ,  $\sigma_2 = 23.9$ , compared to  $\sigma_{\text{residual}} = 15.7$ ). Inclusion of these random slopes was clearly justified by model comparison ( $\chi^2 = 67.8$ ,  $p < .0001$ ), indicating that there was significant variation across participants' quantifier belief distributions.

Although it is well established that context-dependent gradient expectations are ubiquitous in quantifier use, we are not aware of earlier studies that quantify between-talker differences in the usage of quantifiers. The results

establish that there is variation in listeners' expectations of talkers' quantifier usage even when the context is held constant (see also Budescu & Wallsten, 1985, for evidence of between-participant variability in the interpretation of probability terms). This sets the stage for Experiment 2, which asks whether listeners' expectations about how a talker uses quantifiers adapt.

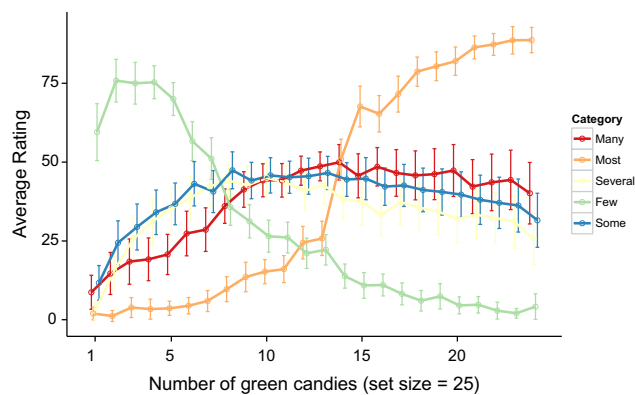
### Experiment 2: Adaptation of beliefs about quantifier use based on recent input

Experiment 2 investigates whether listeners can adjust their beliefs about the use of *some* and *many* based on recent input specific to the current context. We used a variation of the exposure-and-test paradigm frequently used in research on perceptual learning, including research on speech perception (e.g., Eisner & McQueen, 2006; Kraljic & Samuel, 2007; Norris, McQueen, & Cutler, 2003; Vroomen, van Linden, De Gelder, & Bertelson, 2007). A post-exposure test assessed participants' beliefs about the typical use of *some* and *many*. Before this test, participants watched videos of a talker describing various visual scenes with sentences like *Some of the candies are green*. This procedure is illustrated in Fig. 1.

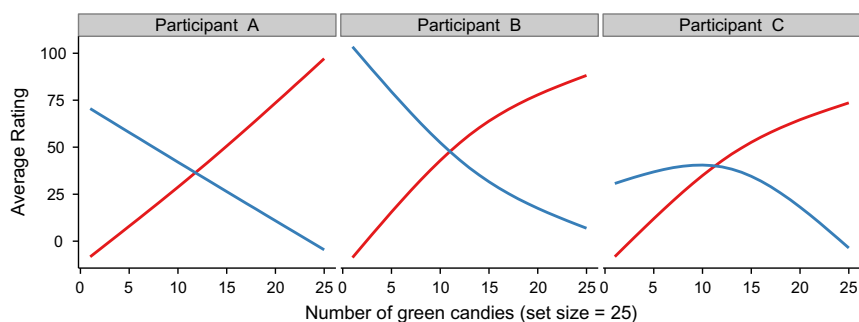
Exposure was manipulated between participants. Half of the participants were exposed to a novel talker's use of the word *some* (some-biased group). Paralleling perceptual recalibration experiments (e.g., Norris et al., 2003), this talker used the quantifier *some* to describe the scene that was maximally ambiguous as to whether it fell in the *some* or the *many* category. This scene (13 green candies, which we refer to as the Maximally Ambiguous Scene or Maximally Ambiguous Scene) was determined on the basis of the ratings from Experiment 1. Using the (fixed effect) parameter estimates from the natural spline fitting procedure described in Experiment 1, we obtained the population-level *some* and *many* curves for all values between 1 and 25. The closest integer to the intersection point of these two curves—i.e. the point that was equally likely to give rise to an expectation for *some* and *many*, 13 green candies—was considered the Maximally Ambiguous Scene. The other half of the participants was exposed to the same novel talker describing the Maximally Ambiguous Scene with the quantifier *many* (many-biased group). This manipulation—with minor modifications—was employed in all experiments reported below.

If passive exposure to a specific talker's use of *many* or *some* is sufficient for listeners to adapt their expectations about the use of *many* and *some*, adaptation should be reflected in shifted belief distributions in the post-test compared to the pre-test. The direction of this shift should depend on the exposure condition. We elaborate on this prediction after introducing the paradigm in more detail below.

As outlined in the introduction, Experiment 2 further aims to assess exactly *what* expectations are affected by exposure to a novel talker's use of *some* and *many*. Specifically, we ask whether exposure to a novel talker leads listeners to develop talker-specific expectations, rather than just changes in expectations that could apply across any type of talker. Episodic (Goldinger, 1996),



**Fig. 2.** Naturalness ratings for the five English quantifiers *few*, *many*, *most*, *several*, and *some*. Error bars are 95% confidence intervals.



**Fig. 3.** Parametric fits to *many* and *some* ratings for three representative participants.

exemplar-based (Johnson, 1997; Pierrehumbert, 2001), certain Bayesian models of adaptation (Kleinschmidt & Jaeger, 2015) predict talker-specific expectations. These models were originally developed to account for adaptation in speech perception, but their logic straightforwardly extends to lexical processing. Indeed, although not necessarily framed as such, memory-based alignment accounts of lexical processing (e.g., Horton & Gerrig, 2005) are essentially exactly such an extension.

To begin to answer whether lexical adaptation to quantifiers can be talker-specific, we conducted two versions of Experiment 2. In Experiment 2a, we asked participants “How likely do you think it is that a speaker will describe this scene with each of these alternatives?”. Using the indefinite “a speaker” leaves ambiguous whether we are referring to a generic talker or to the specific talker they were exposed to. In Experiment 2b, we changed the wording to “How likely do you think it is that the speaker will describe this scene with each of these alternatives?”. Using the definite “the speaker” makes it clear that we are referring to the specific talker they were exposed to. If exposure leads to adaptation globally, i.e., to talkers in general, we expect the adaptation effect to be similar across Experiments 2a and 2b. If instead adaptation is local and specific to the exposure talker (or a mixture of local and global adaptation), there should be a difference in the size of the adaptation effect such that a larger effect should be observed when participants are asked about “the speaker” compared to “a speaker”.

### Experiment 2a

#### Methods

**Participants.** A total of 79 participants were recruited for Experiment 2a via Amazon’s crowdsourcing platform Mechanical Turk. All participants were self-reported native speakers of English. The experiment took about 15 min to complete. Participants were paid \$1.50 (\$6.00/h).

**Materials and procedure.** The experiment proceeded in two phases, illustrated in Fig. 1: the exposure phase (Panel (a)), and the post-exposure test (Panel (b)). The post-exposure test assessed participants’ expectations—the quantifier belief distributions—about talkers’ use of *some* and *many*. Participants saw a bowl of blue and green candies in the center of the scene, and their task was to distribute a sum of 100 points across the three alternative descriptions in response to the question “How likely do you think it is that a speaker will describe this scene with each of these alternatives?”. All participants saw the same set of three alternative explanations: *Some of the candies are green*, *Many of the candies are green*, and “Other”. Assignment of sliders to alternative descriptions was counterbalanced, except that the “Other” alternative was always paired with the right-most slider.

To assess participants’ beliefs about talkers’ use of *some* and *many*, we sub-sampled scenes representing the entire scale. Specifically, a scene could contain one of following number of green candies out of 25: {1, 3, 6, 9, 11, 12, 13,

14, 15, 17, 20, 23}. Over 39 test trials, participants rated each possible number of green candies 3 times. Different instances of the scenes with the same number of green candies differed in the spatial configuration of the blue and green candies. The order of the scenes was pseudo-randomized.

On an exposure phase trial, participants saw a video (Fig. 1(a) illustrates a snap-shot of one such video). We recorded utterances from two talkers (one male and one female), and randomly assigned half of the participants in each of the two groups to one of the two talkers each. The video showed a bowl of 25 candies embedded in the bottom right corner of the video frame. As on post-exposure trials, the bowl always contained a mixture of green and blue candies, but the number and spatial configuration of the candies differed between trials. The video showed a talker describing that scene in a single sentence. The videos played automatically at the start of the trial, and the scene remained visible even when the video finished playing. Participants clicked the “Next” button to proceed. The “Next” button was invisible until the video finished playing to ensure that participants could not skip a video.

Exposure consisted of 10 critical and 10 filler trials. On critical trials, participants saw the Maximally Ambiguous Scene being described by the talker as *Some of the candies are green* (some-biased group) or *Many of the candies are green* (many-biased group). On filler trials, participants observed the talker correctly describing a scene with no green candies as *None of the candies are green* (5 trials) and a scene with no blue candies as *All of the candies are green* (5 trials). Filler trials were included to: (a) make the manipulation less obvious; and (b) encourage participants to believe that the talker was indeed intending to accurately describe the scene. The order of critical and filler trials was pseudo-randomized. Following exposure, participants entered the post-exposure test.

**Catch trials.** In Experiment 2a and in the following experiments, both phases of the experiment contained catch trials after every 2–7 (mean = 5) trials, totaling 15 catch trials (6 during exposure and 9 during post-exposure). Catch trial occurrence was randomized so as to rule out strategic allocation of attention. In this experiment, we excluded one participant whose accuracy was below 75% on the catch trials.

### Predictions

Exposure to a many-biased talker should lead participants to change their beliefs about how this talker uses *many*. These changes could include (i) shifting the many category mean towards the center of the scale (i.e., towards the Maximally Ambiguous Scene), (ii) broadening the many category to include more scenes (towards the Maximally Ambiguous Scene), (iii) increasing the overall probability attributed to that category, or (iv) all of (i)–(iii). Mutatis mutandis, the same predictions hold for the some-biased condition.

Changed beliefs about the biased category (e.g., *many* in the many-biased condition) should also affect participants’ beliefs about how talkers use the alternative lexical categories (e.g., *some* and “Other” in the many-biased condition). Specifically, since participants distributed a fixed number of points across the three alternatives, increased ratings for, e.g., *many* will necessarily affect the other two alternatives (i.e., *some* and “Other”). Given the nature of the exposure phase, which focuses on *many* and *some*, we predict that the trade-off between lexical categories will mostly involve the two quantifiers, rather than the “Other” response.

### Results

Fig. 4(a) illustrates participants’ mean ratings (with 95% confidence intervals) for *many* and *some* in the post-exposure phase. The solid lines show the many-biased group results, and the dotted lines show the some-biased group results. Some-biased participants became more likely to expect *some* to refer to the scenes at the expense of *many*; whereas the many-biased participants became more likely to expect *many* to refer to scenes at the expense of *some* (see Fig. 4(a)).

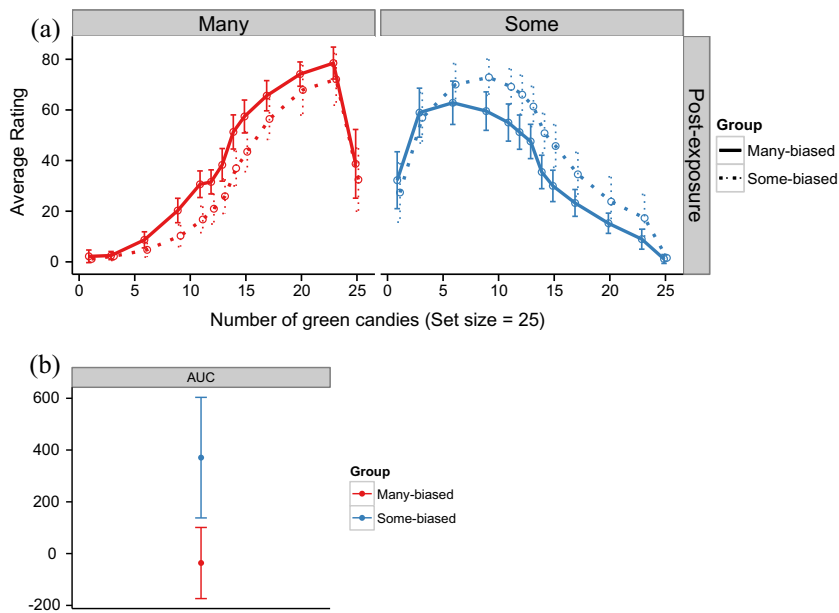
To test whether there indeed was significant adaptation, we quantified the differences between the some-biased and the many-biased groups as the difference in the area under the curve (AUC) in the post-exposure test. We explain this technique in the following.

**Area under the curve.** We fit the category distributions separately for each participant. Based on the mean ratings of a participant, we fit linear models with natural splines with 2 degrees of freedom (Harrell, 2014) independently to the post-exposure *many* ratings and post-exposure *some* ratings. All analyses were conducted using the R statistics software package (R Core Team, 2014). This yielded separate *many* and *some* category distributions for each participant. Compared to the approach taken in Experiment 1, fitting splines separately to each participant does not assume that differences between participants are normally distributed (though that approach yields the same results).

We then calculated the AUC for the *many* curve and the AUC for the *some* curve. Next, we determined each participant’s AUC difference by subtracting the AUC for the *some* category from the AUC for the *many* category (we could have conducted the reverse subtraction, which would only change the direction of effects). The resulting difference score is predicted to be higher for the some-biased group than for the many-biased group, if there is adaptation. As shown in Fig. 4(b), this prediction was borne out ( $t(61.6) = 3.0$ ,  $p < .01$ , two-tailed  $t$ -test allowing for unequal variances).

### Discussion

The results of Experiment 2a demonstrate that just 10 informative critical exposure trials (out of 20 exposure trials) are sufficient to induce lexical adaptation: participants adjusted their beliefs about the use of *many* and *some*



**Fig. 4.** Results for Experiment 2a. (a) Mean ratings in Experiment 2a by participants in the many-biased condition (solid lines) and the some-biased condition (dotted lines). (b) Area under the curve (AUC) analysis. Error bars indicate the 95% confidence intervals.

based on recent input. Experiment 2a used a paradigm closely modeled on the basis of perceptual learning studies.<sup>1</sup>

Recall that in the post-exposure assessment of their beliefs about quantifier use, participants were asked “How likely do you think it is that a speaker will describe this scene with each of these alternatives?” In this question, the referent of “a speaker” is ambiguous between a generic talker and the specific exposure talker. This raises the question of whether the observed adaptation was a result of participants updating their expectations about the exposure talker’s use of *some* and *many*, or whether expectations changed about talkers in general. In Experiment 2b, we used the definite noun phrase “the speaker” to emphasize to participants that we were interested in their beliefs about how the talker they were exposed to would use *many* and *some*.

We present the results of Experiment 2b before further examining the nature of the observed adaptation effect using variants on this basic paradigm in the remainder of the paper.

#### Experiment 2b: Talker- vs. experiment-specific adaptation

Experiment 2b assessed whether adaptation occurred to the specific exposure talker, or whether more general expectations about quantifier use changed.

<sup>1</sup> We replicated this result in a separate experiment not reported here. This experiment was identical to Experiment 2a, but added a pre-exposure test that was identical to the post-exposure test of Experiment 2a. This also allowed us to directly compare changes in the *some* and *many* ratings from the pre- to post-exposure test between the two bias conditions. We found that some-biased exposure resulted in an expansion of the *some* category (and vice versa for the many-biased group). These changes were highly significant, providing a conceptual replication of Experiment 2a.

#### Methods

**Participants.** We recruited 64 participants via Mechanical Turk. The duration and payment were identical to that of Experiment 2a. Three participants were excluded based on their catch trial performance.

**Materials and procedure.** Materials and the procedure for the exposure phase were identical to that of Experiment 2a but the post-exposure test differed in the following two ways. First, each of the alternative descriptions (*Many of the candies are green* and *Some of the candies are green*) were paired with the identical picture of the talker, as shown in Fig. 5. Second, the instructions given prior to the post-exposure test were re-worded so that the indefinite “a speaker” was replaced by the definite “the speaker” in order to emphasize that it is expectations about the specific exposure talker’s likely utterances that are of interest. That is, instead of being asked “How likely do you think it is that a speaker will describe this scene with each of these sentences?” participants were now asked “How likely do you think it is that the speaker will describe this scene with each of these sentences?” (see Fig. 5).

#### Results

Fig. 6(a) illustrates participants’ mean ratings (with 95% confidence intervals) for *many* and *some*. The solid lines show the many-biased group results, and the dotted lines show the some-biased group results. We qualitatively replicate the adaptation effect from Experiment 2a: some-biased participants became more likely to expect *some* to refer to the scenes at the expense of *many*; whereas the many-biased participants became more likely to expect *many* to refer to scenes at the expense of *some*. This difference was significant ( $t(43.8) = 5.4, p < .0001$ ,



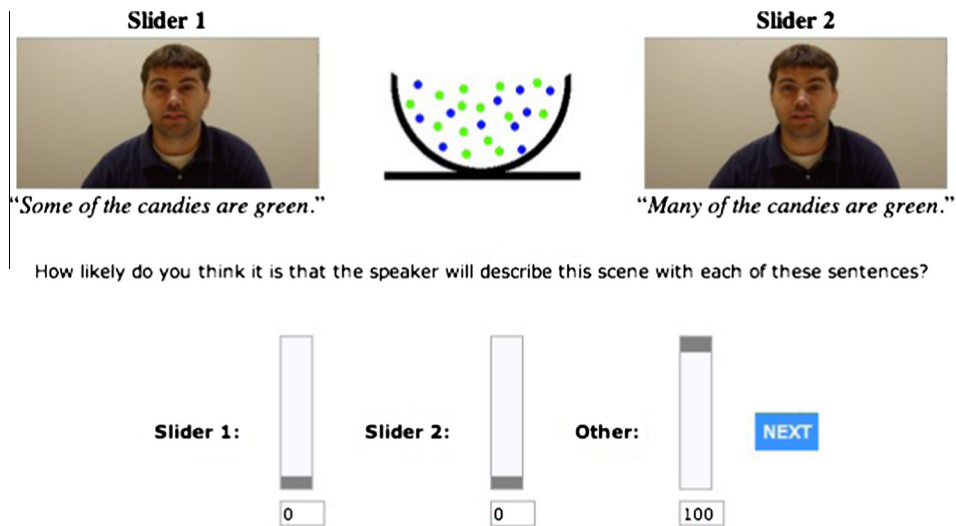


Fig. 5. Snapshot of a post-exposure test trial in Experiment 2b.

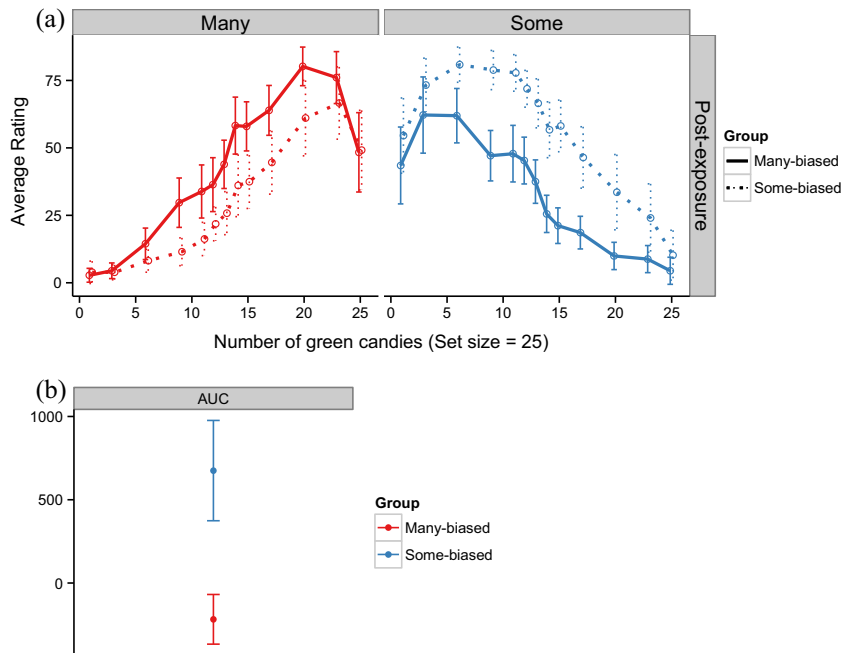


Fig. 6. Results for Experiment 2b. (a) Mean ratings in Experiment 2b by participants in the many-biased condition (solid lines) and the some-biased condition (dotted lines). “Other” ratings are not shown. (b) The area under the curve analysis. Error bars indicate the 95% confidence intervals.

two-tailed *t*-test allowing for unequal variances). Results are shown in Fig. 6(b).

#### Comparison of Experiments 2a and 2b: Talker-specific adaptation

The main question of interest in this experiment was whether the observed adaptation effect is specific to the exposure speaker or generalizes to generic speakers. If some of the participants in Experiments 2a interpreted the instructions to be about a generic talker, we would

expect less transfer from exposure in the post-tests compared to Experiment 2b, where instructions emphasized that the post-test judgments are about the exposure talker. We can directly test this prediction by comparing the results of Experiment 2a to the results of Experiment 2b.

We compared the effect size in Experiments 2a and 2b by calculating Cohen's *D*, a measure of effect size, for both Experiments. Cohen's *D* increased from Experiment 2a (0.7) to Experiment 2b (1.4). We also regressed people's responses against a full factorial of bias (some- vs.

many-bias) and talker-specificity (Experiment 2b: same talker in exposure and test vs. Experiment 2a: generic talker during test). Replicating the *t*-tests reported for Experiments 2a and 2b, there was a main effect of bias ( $\beta = 324.9$ ,  $t = 6.2$ ,  $p < .0001$ ). There was no main effect of talker-specificity ( $ps > .5$ ). Crucially, the interaction between bias and talker-specificity was significant, in that the bias effect was larger when the test speaker was explicitly the same as the exposure speaker (Experiment 2a, compared to Experiment 2b;  $\beta = 121.4$ ,  $t = 2.3$ ,  $p < .05$ ).

This suggests that at least some participants either took Experiment 2a to be about generalization to a generic talker or maintained uncertainty about whether they were asked to generalize to the specific exposure talker or to a generic talker. The comparison of Experiments 2a and 2b further suggests that exposure to the quantifier use of a specific talker primarily leads to adapted expectation about that listener: when we removed uncertainty about whether participants were asked about the specific exposure talker (Experiment 2b), the adaptation effect (measured here as Cohen's *D*) doubled. Taken together, Experiments 2a and 2b thus provide evidence that listeners store lexical experiences along with information about the context in which they occur and that this context includes the talker (Brown-Schmidt et al., 2015; Horton & Gerrig, 2005). This result extends previous evidence for talker-specific lexical expectations from open class words (e.g., Creel et al., 2008; Metzger & Brennan, 2003; also Walker & Hay, 2011). These results are expected under episodic, exemplar-based, and other memory-based alignment accounts (Brown-Schmidt et al., 2015; Horton & Gerrig, 2005; Johnson, 1997; Pierrehumbert, 2001) as well as the rational adaptor account (Kleinschmidt & Jaeger, 2015), but not lexical priming accounts.

### Experiment 3: Adapting to the frequency vs. use of quantifiers

Experiment 3 employed the identical procedure as Experiment 2b, with only one change: participants were exposed to an equal number of *many* and *some* trials. Specifically, the exposure talker produced one of the quantifiers in its prototypical usage (based on Experiment 1 results and confirmed below). For the other quantifier, the exposure talker had the same 'biased' usage employed in Experiment 2. We describe this manipulation in more detail below.

By equating the frequency of *many* and *some* during exposure, Experiment 3 allows us to address whether the adaptation effects observed thus far reflect adaptation of beliefs about the frequency of a particular quantifier (the prior of a quantifier expression), or adaptation of beliefs about the way a given quantifier is used (the likelihood of a quantifier conditional on a set size), or a combination of the two.

There is evidence that listeners can adapt their prior expectations about the frequency with which a given talker uses a word (Creel et al., 2008). A priori, we would expect listeners to be able to adapt their beliefs about both the prior and the likelihood, as both of them are critical in

making robust inferences about the intended meaning. If the source of adaptation is only updated prior expectations of *many* and *some*, Experiment 3 should yield smaller adaptation effects, compared to Experiments 2. If, on the other hand, the sole source of adaptation in Experiments 2 was to changes in the way *many* and *some* are used, then Experiment 3 should continue to yield the same magnitude of adaptation effects observed in Experiments 2. Finally, if the source of adaptation consists of both the updated prior expectations and adaptation to the usage, then Experiment 3 should still yield adaptation, but to a lesser extent than Experiments 2.

### Methods

#### Participants

We recruited 71 participants via Mechanical Turk. The experiment took about 20 min to complete. Participants were paid \$2.00 (\$6.00/h). One participant was excluded due to catch trial performance.

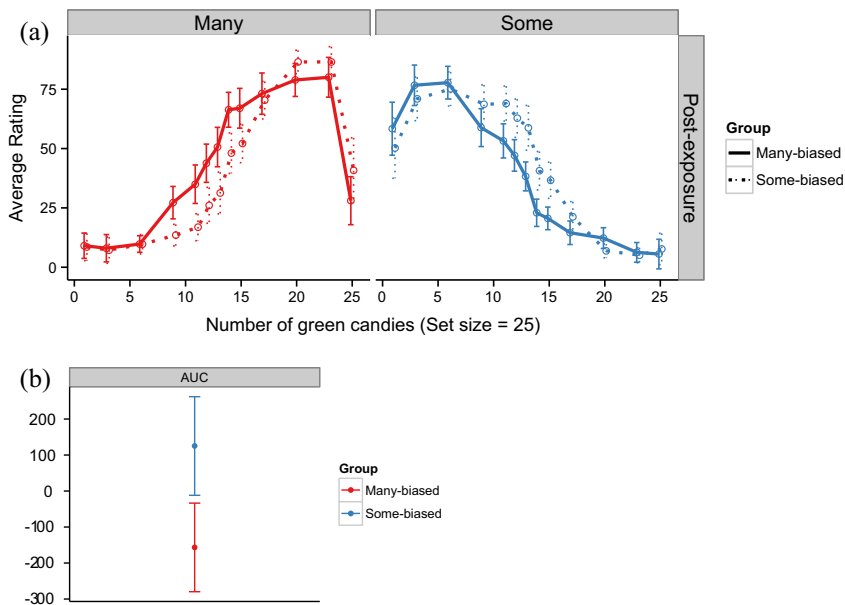
#### Materials and procedures

The procedure was identical to that of Experiment 2b with the exception of additional exposure phase trials. Participants saw a total of 30 exposure trials. Twenty of these trials were identical to those in Experiment 2a. The additional 10 trials exposed participants to a highly typical usage of the other quantifier (typical uses of *many* for the some-biased group and typical uses of *some* for the many-biased group).

The typical *many* trials were generated in the following way. Based on Experiment 1, we selected the scenes with the highest ratings – i.e., the mode and its neighbors – for *many* (the scenes with 20 and 23 green candies) and for *some* (the scenes with 3, 6, and 9 green candies) on the basis of the *some* vs. *many* list results. We did not include the scenes with 25 (i.e., only) green candies for *many* (the other neighbor of the mode for *many*), because the ratings for *many* dropped sharply for that scene. The 10 typical *many* trials were then obtained by embedding scenes with 23 green candies (in 5 of the trials) or with 20 green candies (in the remaining 5 trials) to a talker's video saying *Many of the candies are green*. Likewise, the 10 typical *some* trials were obtained by embedding scenes with 3 green candies (in 3 of the trials), 6 green candies (in 5 of the trials), or 9 green candies (in the remaining 2 trials) to a talker's video saying *Some of the candies are green*.

### Results and discussion

Fig. 7(a) shows the mean post-exposure ratings (error bars indicating the 95% confidence intervals), suggesting an effect in the predicted direction: some-biased participants were more likely to expect *some* to refer to the scenes at the expense of *many*; whereas the many-biased participants were more likely to expect *many* to refer to scenes at the expense of *some*. The difference was significant ( $t(66.8) = 3.1$ ,  $p < .01$ , two-tailed *t*-test allowing for unequal variances). Quantified AUC results are visualized in Fig. 7(b).



**Fig. 7.** Results for Experiment 3. (a) Mean ratings in Experiment 3 by participants in the many-biased condition (solid lines) and the some-biased condition (dotted lines). “Other” ratings are not shown. (b) The area under the curve analysis. Error bars indicate the 95% confidence intervals.

These results – that listeners show the same type of adaptation effect observed in Experiments 2 demonstrate that listeners are adapting, at least in part, to changes in the likelihood of quantifier use conditional on set size, rather than simply the frequency with which a speaker uses a quantifier.

This leaves open whether the adaptation is only of the likelihood, or a mixture of prior and likelihood adaptation. We can begin to address this question by comparing the magnitude of adaptation in Experiments 2b and 3. If the adaptation effect size is not distinguishable across the two experiments, this suggests that the adaptation occurs only in the likelihood. In contrast, a smaller adaptation effect in Experiment 3 than in Experiment 2b would provide evidence for a mixture of prior and likelihood adaptation.

To assess change in effect size, we computed Cohen's  $D$ , a measure of effect size, for both Experiments. Cohen's  $D$  decreased from Experiment 2b (Cohen's  $D$  1.4) to Experiment 3 (Cohen's  $D$  0.7). To test whether this change was significant, we regressed AUC results against the full factorial design of bias (some- vs. many-bias) and experiment (Experiment 2b: repeated exposure to only shifted quantifier use vs. Experiment 3: equi-frequent exposure to both quantifiers). Replicating the  $t$ -tests reported for Experiments 2b and 3, there was a main effect of bias ( $\beta = 293.6$ ,  $t = 6.4$ ,  $p < .0001$ ). This effect interacted significantly with experiment, in that it was smaller when participants saw both quantifiers equally often (Experiment 3;  $\beta = 152.7$ ,  $t = 3.3$ ,  $p < .01$ ). This suggests that changes in participants' beliefs about the prior frequency of *many* and *some* contribute to the adaptation effects observed in Experiments 2. At the same time, Experiment 3 extends previous research on lexical adaptation (Creel et al.,

2008; Metzing & Brennan, 2003): To the best of our knowledge, Experiment 3 is the first to suggest that listeners can adapt their beliefs about *how* (i.e., with what intended interpretation) specific talkers use a given word, in our case *many* and *some*. We believe that research addressing this question will be particularly important for further quantitative investigations and modeling.<sup>2</sup>

There was also a main effect of experiment (Experiment 3;  $\beta = 122.1$ ,  $t = 2.7$ ,  $p < .01$ ). We had no specific expectations about this main effect, but it could point to asymmetries in the strengths of prior beliefs about the typical distribution of *many* and *some* (specifically, asymmetries in the beliefs about how the use of the two quantifiers differ across talkers). The main effect could also point to prior beliefs about how the type of exposure talker we used in our experiments differs from generic talkers (e.g., based on how they were dressed in the exposure video, their speech style or dialectal background). We leave these questions to future research.

<sup>2</sup> For example, we note that even under very general assumptions about adaptation, exposure to equally many typical uses of *many* and *some* is not sufficient to completely rule out adaptation of beliefs about the prior frequencies of the two quantifiers. One reason for this is that exposure to equally many trials is expected to affect a priori less frequent quantifier more strongly (cf. the ideal adapter framework presented in Kleinschmidt & Jaeger, 2015). Indeed, *many* and *some* are not equally frequent in participants' previous experience. For example, corpus counts indicate that the bigram *some of* occurs at least twice as often as *many of* (27,601 vs. 12,919 occurrences in the British National Corpus, respectively). However, adaptation of only the prior seems an unlikely explanation for the current results, given that the AUC change in the many- vs. some-biased conditions was overall symmetrical around 0 (see Figs. 6 and 7) – thus suggesting relatively similar changes to the representations of *many* and *some*.

#### Experiment 4: Adapting to multiple talkers

Taken together, Experiments 2 and 3 suggest that exposure to relatively few trials is sufficient for listeners to adapt their expectations about how a given talker uses *many* and *some* – at least, when the talker is observed producing highly informative descriptions as in the current experiments, where the talker is observed producing multiple critical trials describing the same domain.

As we noted earlier, the need for adaptation and the need to maintain previously acquired knowledge can be balanced by learning and maintaining talker-specific expectations. Experiment 4 investigates whether listeners can adapt to the lexical preferences of multiple talkers simultaneously. Building on the paradigm used in Experiment 2b, participants observed the lexical preferences of two different talkers in a blocked exposure phase (e.g., exposure to a some-biased talker, followed by exposure to a many-biased talker). Participants then rated descriptions by each talker in a blocked post-exposure test. If participants adapt their expectations of quantifier use in a talker-specific manner, we should observe adaptation effects in opposite directions for the two speakers.

#### Methods

##### Participants

We recruited 54 participants via Mechanical Turk. The experiment took about 25 min to complete. Participants were paid \$2.50 (\$6.00/h). Two participants were excluded because of their catch trial performance.

##### Materials and procedure

Unlike in Experiments 2 and 3, there were two exposure blocks and two post-exposure test blocks. Each exposure block featured a different talker. Each post-exposure test block tested for one of the exposure talkers. The two talkers now used as a within-participant manipulation were the same male and female talker used in Experiments 2 and 3 in the between-participant designs.

Materials and the procedure for each pair of exposure and post-exposure test blocks (i.e., blocks playing and testing the same talker) were identical to that of Experiment 2b (see Fig. 5 above). One of the exposure talkers was many-biased. The other one was some-biased. Both exposure blocks preceded both post-exposure test blocks. For example, a participant might see an exposure block with the many-biased male talker, followed by an exposure block with the some-biased female talker, followed by a post-exposure test block for the male talker, and finally a post-exposure test block for the female talker.

Across participants, we counter-balanced (a) the order of talker-gender in the exposure blocks (male talker first vs. female talker first), (b) the order of talker-bias in the exposure block (many-biased first vs. some-biased first), which also balanced the talker-bias to talker-gender assignment (whether the male or the female talker was many-biased and, hence, whether the male or female talker was some-biased), and (c) whether the order of post-exposure test blocks was the same or inverse of the

order of exposure blocks. All eight factorial combinations of these  $2 \times 2 \times 2$  nuisance variables occurred equally often across participants.

#### Results and discussion

We first analyze the overall adaptation effect across all orders of talker-gender and talker-bias using the same analysis as in Experiments 2 and 3. After establishing that the basic adaptation effect observed in Experiments 2 and 3 is also observed when two exposure speakers are used, we assess whether these talker-specific adaptation effects were affected by the order of presentation (e.g., due to recency or interference effects). Such order effects would begin to point to some of the limits of talker-specific lexical adaptation.

##### Overall adaptation effect

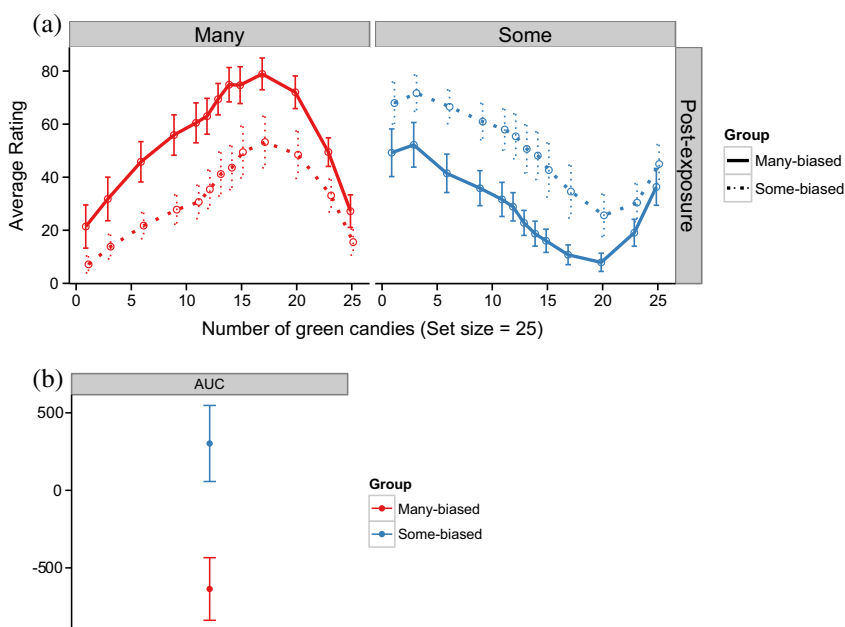
Fig. 8(a) illustrates mean post-test ratings for *many* and *some*, collapsing over all orders of talker-bias and talker-gender in exposure and post-test. Participants' expectations about quantifier use adapted in response to exposure in the predicted direction: when tested for some-biased speakers, listeners were more likely to expect *some* to refer to the scenes at the expense of *many*; whereas when tested for the many-biased speakers, they were more likely to expect *many* to refer to scenes at the expense of *some*. Importantly, they did so even though there were two different talkers with different preferences in the way that they used quantifiers, indicating that participants separately adapted to each talker's preference. We quantified the overall adaptation again collapsing data from both post-exposure test blocks across all participants. The results shown in Fig. 8(b) indicate that the listeners tracked each talker's lexical preferences and adapted their interpretations accordingly ( $t(98.3) = 5.9$ ,  $p < .0001$ , two-tailed  $t$ -tests allowing for unequal variances).

Thus participants show talker-specific adaptation to two talkers. This leaves open whether adaptation to multiple talkers in any way reduces the individual adaptation results. For example, it is possible that the experiences with the two talkers interfere with one another in memory or that more recent exposure overrides less recent experience. To assess these questions, we conducted a regression analysis.

##### Analysis of order effects

Using linear regression, we regressed AUC values against the factorial design of talker-bias (some- vs. many-bias), talker-exposure order (1st vs. 2nd, i.e., whether the current post-test talker was seen during the 1st or the 2nd exposure block), and talker-test order (1st vs. 2nd; i.e., whether the current post-exposure test block is the 1st or the 2nd). If exposure to the two different speakers leads to interference with each other, the effect of the talker-bias should interact with talker-exposure order, talker-test order, or their interaction.

There was a main effect of talker-bias ( $\beta = 467.8$ ,  $t = 5.8$ ,  $p < .0001$ ), paralleling our  $t$ -test above. There was no main effect of any of the other variables or their interactions ( $ps > .25$ ).



**Fig. 8.** Results for Experiment 4. (a) Mean ratings in Experiment 4 by participants in the many-biased condition (solid lines) and the some-biased condition (dotted lines). “Other” ratings are not shown. (b) The area under the curve analysis. Error bars indicate the 95% confidence intervals.

These results suggest that listeners can develop talker-specific expectations about quantifier use for at least two talkers. The regression analysis did not reveal evidence of recency or interference effects, consistent with similar findings in phonetic adaptation (e.g., Kraljic & Samuel, 2007). This suggests that, at least when provided with highly informative signals about talker-specific preferences in quantifier use, listeners can readily adapt to two talkers. We note, however, that Experiment 4 might not have sufficient power to detect relatively subtle interference effects: while talker-bias was manipulated within-participants, talker-exposure and talker-test order were manipulated between-participants, reducing the power to detect effects of these factors or their interaction with talker-bias. We thus consider it an open question for future work, whether or to what extent adaptation to talker-specific quantifier use decays over time or interferes with adaptation to other talkers.

## General discussion

The studies reported in this paper used a web-based paradigm to explore a specific type of context-dependence that has received comparatively little attention in the literature, talker-specific differences in how quantifiers are used. In particular we focused on adaptation to talker-specific use of *some* and *many*, drawing parallels to recent work on adaptation in other domains of language, with a special focus on phonetic adaptation—the domain, which has been most widely investigated to date. Experiment 1 demonstrated that listeners vary in their expectations for how a given talker will use quantifiers. This establishes that adaptation would be useful for

efficient communication. Experiment 2a used an exposure, post-exposure test design, modeled on work in perceptual learning for phonetic categories, finding that listeners who were exposed to a speaker who used *some* to describe the most ambiguous scene (13 of 25 candies) exhibited a different quantifier belief distribution than listeners exposed to a speaker who used *many* to describe the most ambiguous scene. Experiment 2b used a variation on the paradigm used in Experiment 2a to establish that listeners were primarily adapting to the specific talker they were exposed to, rather than a generic talker. Experiment 3 demonstrated that adaptation occurred even when the frequency of quantifier use in the exposure phase was equated. Comparisons of the effect sizes in Experiments 2b and 3 demonstrated that listeners were adapting both to the frequency of quantifier use by a talker and the likelihood of quantifier use for a particular scene. Finally Experiment 4 demonstrated that listeners learned and maintained expectations about the quantifier use for two different talkers.

In the remainder of this section, we discuss the implications that this work has for the role of adaptation in language use and the modeling frameworks that are likely to have the capability of capturing the reported data.

Taken together, the results of Experiments 2 through 4 demonstrate that, based on brief exposure, listeners update their expectations about how a talker will use quantifiers to refer to entities in simple displays. These findings contribute to a growing body of work suggesting that listeners rapidly adapt to talker-specific information at multiple linguistic levels, including phonetic categorization, use of prosody, lexical choice, and use of syntactic structures (e.g., Creel & Bregman, 2011; Fine & Jaeger, 2013; Kamide, 2012; Kraljic & Samuel, 2007; Kurumada



et al., 2014; Norris et al., 2003). This work implicates adaptation as a fundamental process by which listeners cope with the well-documented variability in language use both between and within talkers. For example, talker-specific information affects spoken word recognition (Creel & Bregman, 2011; Creel & Tumlin, 2009; Goldinger, 1998) as well as listeners' expectations about a specific talker's use of concrete nouns to refer to entities in a scene (Creel et al., 2008). More recent studies also suggest that talker-specific expectations are even observed during syntactic processing (Kamide, 2012). The current studies build upon this work by extending it to quantifier use and interpretation.

One consequence of these studies, including the current experiments, is that standard, non-strategic, priming accounts of alignment – be it in speech perception, lexical, or syntactic processing (e.g., Goudbeek & Krahmer, 2012; Pickering & Garrod, 2004; Reitter et al., 2011; Traxler & Tooley, 2008)—are insufficient to account for existing data. This includes even some implicit learning accounts of priming (Chang et al., 2006). None of these accounts predicts talker-specific expectations (for related discussion of lexical priming accounts, see also Heller & Chambers, 2014). Instead, the current results – in particular, Experiments 2 and 4—provide further evidence that listeners can learn and store (at least for the period of an experiment) linguistic experiences along with rich knowledge about the context these experienced occurred in. This result is in line with memory-based accounts (Goldinger, 1996; Johnson, 1997; Pierrehumbert, 2001) that consider talkers to be part of the contextual information that linguistic experiences are stored with (for further discussion, see Brown-Schmidt et al., 2015; Kleinschmidt & Jaeger, 2015). The memory processes evoked by these accounts are taken to be typically automatic and implicit (see also Horton & Gerrig, 2005). An interesting question for future research is whether the same type of memory-based explanations can also explain those priming effects that are often taken to require separate explanations in terms of “non-strategic” processes (Pickering & Ferreira, 2008; Traxler & Tooley, 2008).

Second, the current work establishes a foundation for future empirical and computational investigations of how listeners interpret quantifiers and other linguistic expressions with abstract meanings. It will be important to establish the conditions under which listeners adapt. For example, one important set of open questions concerns generalization. Quantifier use varies with context. For instance, compare an utterance such as *Bill has many cars* to *Bill has many antiques*. It is likely that about three or more cars could qualify as *many cars* whereas it seems that a higher number of antiques would be necessary to qualify as *many antiques* (see Hörmann, 1983, for many other examples). This raises questions about how adaptation to one domain (e.g., cars) generalizes to other domains (e.g., antiques). Two important questions will be the degree to which results obtained with one type of quantity will generalize to another and the degree to which listeners will assume that a talker who, for instance, uses *some* to refer to greater quantities of candies than a typical speaker is

also likely to use *some* to refer to larger quantities in general or only across similar types of domains.

Similar questions arise about talker-specificity and generalization across groups of talkers. Experiments 2b to 4 suggest that lexical adaptation can be talker-specific. However, the comparison between Experiments 2a and 2b leaves open the question of whether adaptation also generalizes beyond the specific talker. Recall that Experiment 2a left it to participants whether they took the post-exposure test to be about the specific exposure talker or talkers in general, whereas Experiment 2b unambiguously asked about the specific exposure talker. We observed significantly stronger adaptation effects in Experiment 2b (i.e., when listeners were asked about the specific talker they were exposed to). However, we also observed adaptation when participants were asked about talkers in general (Experiment 2a). This result leaves open the question of whether adaptation *also* generalizes beyond the specific exposure talker. On the one hand, it is possible that exposure to a novel talker *only* affects expectations to that talker. On the other hand, it is possible that such exposure *also* generalizes to expectations about other talkers. We briefly elaborate on these two possibilities, as we take them to be an interesting venue for future research (for related discussion, see also Gorman, Gegg-Harrison, Marsh, & Tanenhaus, 2013).

Generalization is, in fact, explicitly predicted under the accounts cited above (Horton & Gerrig, 2005; Johnson, 1997; Pierrehumbert, 2001; Kleinschmidt & Jaeger, 2015), though for slightly different reasons. In the model proposed by Horton and Gerrig (2005), generalization to other talkers follows because memory is *faulty*. In episodic and exemplar-based accounts, generalization takes place because any experience is assumed to become part of the cloud of stored knowledge that future processing draws on (see also Kleinschmidt & Jaeger, 2015; Weatherholtz & Jaeger, submitted for publication). These accounts also explicitly predict that generalization from an exposure talker to other talkers should be strongest when there is little other previous experience with the broader context in which the exposure talker was experienced (as is arguably the case in Experiment 2a).

In pursuing these and related questions, we believe it will be necessary to take a two-pronged approach, combining behavioral paradigms like the one introduced here with computational models that provide clear quantitative predictions about how listeners integrate prior linguistic experience and recent experience with a specific linguistic environment. Although considerable progress has been made both in the development of the computational frameworks (for recent overviews, see e.g., Clark, 2013; Friston, 2005) and the development of paradigms suitable for the study of incremental adaptation (e.g. Fine & Jaeger, 2013; Vroomen et al., 2007), it is only recently that these two approaches are being integrated. To the best of our knowledge, computational modeling of adaptation behavior has so far mostly been limited to adaptation in speech perception (though see Chang et al., 2006; Fine et al., 2010; Kleinschmidt et al., 2012; Reitter et al., 2011 for models of syntactic adaptation), including Bayesian

models (Kleinschmidt & Jaeger, 2015), connectionist models (Lancia & Winter, 2013; Mirman, McClelland, & Holt, 2006), and exemplar-based approaches (e.g., Johnson, 1997; Pierrehumbert, 2001).<sup>3</sup> Developing and applying these types of related models to the domain of quantifier use will allow for formal tests of hypotheses about the principles that listeners use to generalize word meanings across speakers.

Given the strength of the signal that participants were exposed to, it will be important for future work to explore the limits on adaptation in more naturalistic settings. A further interesting open question is whether the adaptation effects observed here are reflected in online language understanding. Recent research on adaptation during speech perception (see, e.g., Trude & Brown-Schmidt, 2012; Creel et al., 2008), syntactic processing (Fine and Jaeger, 2013; Kamide, 2012), and prosodic processing (Kurumada et al., 2014) provides examples of how these questions can be addressed.

## Conclusion

The experiments reported in this paper suggests that even minimal exposure to a speaker whose use of quantifiers differs from a listener's expectations can result in a talker-specific shift in that listener's beliefs about future quantifier use. Our results further suggest that listeners adapt to both the frequency with which a talker uses certain words and the specific interpretation intended by the talker. This complements work on adaptation in other domains, for example, adaptation in response to phonetically or syntactically deviant input and talker-specificity in linguistic processing. The work reported here provides further evidence that listeners can adapt to individual speakers' language use, remember these talker-specific preferences, and use this knowledge to guide utterance interpretation.

## Acknowledgments

We thank Andrew Watts for technical support. We thank the editor and the four anonymous reviewers for their helpful comments on the earlier versions of this manuscript. Parts of this study were presented at the 35th Annual Conference of the Cognitive Science Society, the 20th Architectures and Mechanisms for Language Processing, and at *Discourse Expectations: Theoretical, Experimental, and Computational Perspectives* 2015. This work was partially supported by a post-doctoral fellowship to IY (through the Center for Brains, Minds, and Machines, funded by NSF STC award CCF-1231216), by an SNSF Early Postdoc. Mobility fellowship to JD, by NSF CAREER award IIS-1150028 as well as an Alfred P. Sloan Research Fellowship to TFJ, and by NIH grant HD 27206 to MKT. The views expressed here are those of the authors and do not necessarily reflect those of any of these funding agencies.

<sup>3</sup> Kleinschmidt and Jaeger (2015) discusses many of these different models and the extent to which they capture existing data on talker-specificity, adaptation, and generalization in speech perception.

## References

- Allen, J. S., Miller, J. L., & DeSteno, D. (2003). Individual talker differences in voice-onset-time. *Journal of the Acoustical Society of America*, 113, 544–552.
- Arai, M., van Gompel, R. P. G., & Scheepers, C. (2007). Priming ditransitive structures in comprehension. *Cognitive Psychology*, 54, 218–250. <http://dx.doi.org/10.1016/j.cogpsych.2006.07.001>.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390–412.
- Bach, K. (2012). Saying, meaning, and implicating. In K. Allan & K. Jaszczolt (Eds.), *The Cambridge handbook of pragmatics*. Cambridge University Press.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). *lme4: Linear mixed-effects models using Eigen and S4*. Retrieved from <<http://cran.r-project.org/package=lme4>>.
- Bauer, L. (1985). Tracing phonetic change in the received pronunciation of British English. *Journal of Phonetics*, 13, 61–81. Retrieved from <<http://cat.inist.fr/?aModele=afficheN&cpsidt=8639296>>.
- Branigan, H. P., Pickering, M. J., & McLean, J. F. (2005). Priming prepositional-phrase attachment during comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 468.
- Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 1482–1493. Retrieved from <<http://www.ncbi.nlm.nih.gov/pubmed/8921603>>.
- Brown-Schmidt, S., Yoon, S. O., & Ryskin, R. A. (2015). Chapter three – People as contexts in conversation. *Psychology of Learning and Motivation*, 62, 59–99. <http://dx.doi.org/10.1016/bs.plm.2014.09.003>.
- Budescu, D. V., & Wallsten, T. S. (1985). Consistency in interpretation of probabilistic phrases. *Organizational Behavior and Human Decision Processes*, 36, 391–405.
- Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, 113, 234.
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of Cognitive Science. *Behavioral and Brain Sciences*, 36, 181–204.
- 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. <http://dx.doi.org/10.1016/j.cognition.2008.04.004>.
- Creel, S. C., & Tumlin, M. A. (2009). Talker information is not normalized in fluent speech: Evidence from on-line processing of spoken words. In *31st Annual conference of the cognitive science society*. Amsterdam, The Netherlands.
- Creel, S. C., Aslin, R. N., & Tanenhaus, M. K. (2008). Heeding the voice of experience: The role of talker variation in lexical access. *Cognition*, 106, 633–664. <http://dx.doi.org/10.1016/j.cognition.2007.03.013>.
- Creel, S. C., & Bregman, M. R. (2011). How talker identity relates to language processing. *Language and Linguistics Compass*, 5, 190–204. <http://dx.doi.org/10.1111/j.1749-818X.2011.00276.x>.
- Eisner, F., & McQueen, J. M. (2006). Perceptual learning in speech: Stability over time. *The Journal of the Acoustical Society of America*, 119, 1950. <http://dx.doi.org/10.1121/1.2178721>.
- Fine, A. B., & Jaeger, T. F. (2013). Evidence for implicit learning in syntactic comprehension. *Cognitive Science*, 37, 578–591. <http://dx.doi.org/10.1111/cogs.12022>.
- Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PLoS One*, 8, e77661. <http://dx.doi.org/10.1371/journal.pone.0077661>.
- Fine, A. B., Qian, T., Jaeger, T. F., & Jacobs, R. A. (2010). Is there syntactic adaptation in language comprehension? In J. T. Hale (Ed.), *Proceedings of ACL: Workshop on cognitive modeling and computational linguistics* (pp. 18–26). Stroudsburg, PA: Association for Computational Linguistics.
- Finegan, E., & Biber, D. (2001). *Register variation and social dialect variation: The register axiom. Style and sociolinguistic variation*. Cambridge University Press.
- Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, 360, 815–836. <http://dx.doi.org/10.1098/rstb.2005.1622>.
- Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 1166.
- Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105, 251.
- Gorman, K. S., Gegg-Harrison, W., Marsh, C. R., & Tanenhaus, M. K. (2013). What's learned together stays together: Speakers' choice of referring

- expression reflects shared experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39, 843.
- Goudbeek, M., & Krahmer, E. (2012). Alignment in interactive reference production: Content planning, modifier ordering, and referential overspecification. *Topics in Cognitive Science*, 4, 269–289.
- Halff, H. M., Ortony, A., & Anderson, R. C. (1976). A context-sensitive representation of word meanings. *Memory & Cognition*, 4, 378–383. <http://dx.doi.org/10.3758/BF03213193>.
- Harrell, F. E. J. (2014). *rms: Regression modeling strategies*. Retrieved from <<http://cran.r-project.org/package=rms>>.
- Harrington, J., Palethorpe, S., & Watson, C. (2000). Monophthongal vowel changes in Received Pronunciation: An acoustic analysis of the Queen's Christmas broadcasts. *Journal of the International Phonetic Association*, 30, 63–78.
- Heller, D., & Chambers, C. G. (2014). Would a blue kite by any other name be just as blue? Effects of descriptive choices on subsequent referential behavior. *Journal of Memory and Language*, 70, 53–67.
- Hörmann, H. (1983). *Was tun die Wörter miteinander im Satz? oder wieviele sind einige, mehrere und ein paar?* Göttingen: Verlag für Psychologie, Dr. C.J. Hogrefe.
- Horton, W. S., & Gerrig, R. J. (in press). Revisiting the memory-based processing approach to common ground. *Topics in Cognitive Science*.
- Horton, W. S., & Gerrig, R. J. (2005). The impact of memory demands on audience design during language production. *Cognition*, 96, 127–142. <http://dx.doi.org/10.1016/j.cognition.2004.07.001>.
- Jaeger, T. F., & Snider, N. E. (2013). Alignment as a consequence of expectation adaptation: Syntactic priming is affected by the prime's prediction error given both prior and recent experience. *Cognition*, 127, 57–83. <http://dx.doi.org/10.1016/j.cognition.2012.10.013>.
- Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics*, 34, 485–499.
- Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In K. Johnson & J. Mullenix (Eds.), *Talker variability in speech processing* (pp. 145–166). Academic Press.
- Kamide, Y. (2012). Learning individual talkers' structural preferences. *Cognition*, 124, 66–71.
- Kamp, H. (1995). Prototype theory and compositionality. *Cognition*, 57, 129–191. [http://dx.doi.org/10.1016/0010-0277\(94\)00659-9](http://dx.doi.org/10.1016/0010-0277(94)00659-9).
- Kennedy, C., & McNally, L. (2005). Scale structure, degree modification, and the semantics of gradable predicates. *Language*, 81, 345–381. <http://dx.doi.org/10.2307/4489896>.
- Klein, E. (1980). A semantics for positive and comparative adjectives. *Linguistics and Philosophy*, 4, 1–45. <http://dx.doi.org/10.1007/BF00351812>.
- Kleinschmidt, D. F., Fine, A. B., & Jaeger, T. F. (2012). A belief-updating model of adaptation and cue combination in syntactic comprehension. In *Proceedings of the 34th Annual Meeting of the Cognitive Science Society (CogSci12)* (pp. 599–604). Austin, TX: Cognitive Science Society.
- Kleinschmidt, D. F., & Jaeger, T. F. (2011). A Bayesian belief updating model of phonetic recalibration and selective adaptation. In *Proceedings of the cognitive modeling and computational linguistics workshop at ACL* (pp. 10–19).
- Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognizing the familiar, generalizing to the similar, and adapting to the novel. *Psychological Review*, 122, 148.
- Kleinschmidt, D. F., & Jaeger, T. F. (in press). Re-examining selective adaptation: Fatiguing feature detectors, or distributional learning? *Psychonomic Bulletin & Review*.
- Kraljic, T., & Samuel, A. G. (2007). Perceptual adjustments to multiple speakers. *Journal of Memory and Language*, 56, 1–15.
- Kurumada, C., Brown, M., & Tanenhaus, M. K. (2012). Pragmatic interpretation of contrastive prosody: It looks like speech adaptation. In *Proceedings of the 34th Annual Meeting of the Cognitive Science Society* (pp. 647–652).
- Kurumada, C., Brown, M., Bibyk, S., Pontillo, D., & Tanenhaus, M. K. (2014). Rapid adaptation in online pragmatic interpretation of contrastive prosody. In *Proceedings of the 37th Annual Meeting of the Cognitive Science Society*.
- Lancia, L., & Winter, B. (2013). The interaction between competition, learning, and habituation dynamics in speech perception. *Laboratory Phonology*, 4, 221–257.
- McCloskey, M., & Cohen, N. J. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. In G. H. Bower (Ed.), *Psychology of learning and motivation*. New York: Academic Press.
- McRae, K., & Hetherington, P. (1993). Catastrophic interference is eliminated in pre-trained nets. In *Proceedings of the 15th Annual Meeting of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.
- Metzing, C., & Brennan, S. E. (2003). When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions. *Journal of Memory and Language*, 49, 201–213. [http://dx.doi.org/10.1016/S0749-596X\(03\)00028-7](http://dx.doi.org/10.1016/S0749-596X(03)00028-7).
- Mirman, D., McClelland, J. L., & Holt, L. L. (2006). An interactive Hebbian model of lexically guided tuning of speech perception. *Psychonomic Bulletin & Review*, 13, 958–965.
- Newstead, S. E. (1988). Quantifiers as fuzzy concepts. In T. Zetenyi (Ed.), *Fuzzy sets in psychology* (pp. 51–72). Amsterdam: Elsevier Science.
- Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, 47, 204–238.
- Pepper, S., & Prytlak, L. S. (1974). Sometimes frequently means seldom: Context effects in the interpretation of quantitative expressions. *Journal of Research in Personality*, 8, 95–101.
- Pickering, M. J., & Branigan, H. P. (1998). The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language*, 39, 633–651.
- Pickering, M. J., & Ferreira, V. S. (2008). Structural priming: A critical review. *Psychological Bulletin*, 134, 427.
- Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27, 169–190.
- Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee & P. Hopper (Eds.), *Frequency effects and the emergence of linguistic structure* (pp. 137–157). Amsterdam: John Benjamins.
- R Core Team (2014). *R: A language and environment for statistical computing*. Vienna, Austria. Retrieved from <<http://www.r-project.org>>.
- Reitter, D., Keller, F., & Moore, J. D. (2011). A computational cognitive model of syntactic priming. *Cognitive Science*, 35, 587–637.
- Roland, D., Dick, F., & Elman, J. L. (2007). Frequency of basic English grammatical structures: A corpus analysis. *Journal of Memory and Language*, 57, 348–379.
- Seidenberg, M. S. (1994). Language and connectionism: The developing interface. *Cognition*, 50, 385–401.
- Tagliamonte, S., & Smith, J. (2005). No momentary fancy! The zero “complementizer” in English dialects. *English Language and Linguistics*, 9, 289–309.
- Traxler, M. J., & Tooley, K. M. (2008). Priming in sentence comprehension: Strategic or syntactic? *Language and Cognitive Processes*, 23, 609–645.
- Trude, A. M., & Brown-Schmidt, S. (2012). Talker-specific perceptual adaptation during on-line speech perception. *Language and Cognitive Processes*, 27, 979–1001.
- Vroomen, J., van Linden, S., De Gelder, B., & Bertelson, P. (2007). Visual recalibration and selective adaptation in auditory-visual speech perception: Contrasting build-up courses. *Neuropsychologia*, 45, 572–577.
- Walker, A., & Hay, J. (2011). Congruence between ‘word age’ and ‘voice age’ facilitates lexical access. *Laboratory Phonology*, 219–237.
- Weatherholtz, K., & Jaeger, T. F. (2015). Speech perception and generalization across talkers and accents [submitted for publication].
- Weiner, E. J., & Labov, W. (1983). Constraints on the agentless passive. *Journal of Linguistics*, 19, 29–58.
- Yaeger-Dror, M. (1994). Phonetic evidence for sound change in Quebec French. *Phonological Structure and Phonetic Form*, 267–293.