Explaining away in semantic-pragmatic adaptation

Sebastian Schuster, Matthew Iver Loder, and Judith Degen

Department of Linguistics, Stanford University

Abstract

Previous work has shown that listeners deal with variability in language use through adaptation; they update expectations about a specific speaker's productions based on the interactions with that speaker. We explore whether contextual information such as a speaker's mood can influence the extent to which listeners adapt to variable use of uncertainty expressions like *might* and *probably*. We find that information about the speaker's mood influences participants' expectations about a generic speaker's use of uncertainty expressions (Exp. 1). We further find that information about the speaker's mood influences adaptation behavior: listeners adapt less to a speaker if they are provided with a reason for the observed language use (Exp. 2), though not more when the behavior is highly unexpected (Exp. 3). These results suggest that listeners explain away otherwise unexpected behavior when presented with a reason, and that the extent of adaptation depends on prior expectations about language use.

Keywords: adaptation; semantics; pragmatics; uncertainty expressions, explaining away

Explaining away in semantic-pragmatic adaptation

Introduction

Variability across speakers can be observed at all levels of linguistic representation and no two speakers use language exactly the same way. For example, at the lexical level, one speaker may use the quantifier *some* to describe a proportion of about 50% whereas another speaker may use the quantifier *many* in that situation (Yildirim, Degen, Tanenhaus, & Jaeger, 2016). Listeners deal with this variability by adapting to individual speakers and forming speaker-specific expectations about language use (Norris, McQueen, & Cutler, 2003; Kraljic & Samuel, 2005; Bradlow & Bent, 2008; Kurumada, Brown, & Tanenhaus, 2012; Kamide, 2012; Kleinschmidt & Jaeger, 2015; Fine & Jaeger, 2016; Roettger & Franke, 2019, inter alia). For example, for utterances with quantifiers, Yildirim et al. (2016) found that this variability was reflected in participants' initial beliefs about the use of quantifiers: participants were uncertain whether a generic speaker would use the utterance *Some of the candies are green* or the utterance *Many of the candies are green* to describe a bowl in which there were approximately equal numbers of green and blue candies. However, if participants briefly observed a speaker describing this scene either consistently using *some* or consistently using *many*, they updated their expectations about how that speaker would use quantifiers to describe different proportions to closely mirror the observed speaker's usage. Similarly, Schuster and Degen (2020) found that listeners update their expectations about a speaker's mapping between uncertainty expressions like *might* or *probably* and event probabilities after brief exposure to that speaker; when exposed to a speaker who consistently used *might* to describe an event probability of 60%, participants expected the speaker to use *might* for higher event probabilities than when when exposed to a speaker who consistently used *probably* to describe the same event probability.

Most investigations into how listeners adapt so far have been based on the assumption that a lot of the variability is *intrinsic* to the speaker and does not vary across contexts. This assumption is reasonable for many kinds of variability. For example, a lot of acoustic properties of productions are determined by a combination of the speaker's gender and the geographical region

that they grew up in (Kleinschmidt, 2019), which are both properties of the speaker that generally do not change between contexts. If one interacts with a woman with a Brooklyn accent, it is highly unlikely that she will sound like an average male speaker with a Southern accent the next time one talks to her. Or, returning to the lexical level, if one interacts with a speaker whose semantic representation of the quantifier *a couple of* is equivalent to the representation of *exactly two*, it is unlikely that they would suddenly use *a couple of* to refer to small quantities greater than two.

At the same time, however, variability in language use can often be attributed to contextual factors other than the speaker, and listeners consider these factors in interpretations. For example, the interpretation of quantifiers and uncertainty expressions depends on the distribution over quantities and base probability rates: *a few mountains* is generally interpreted to refer to a lower quantity than *a few crumbs* (Clark, 1991; Schöller & Franke, 2017) and *probably* describing the likelihood of snow in the North Carolina mountains in December is generally interpreted to describe higher event probabilities than when *probably* is used to indicate the likelihood of snow in October (Wallsten, Fillenbaum, & Cox, 1986). Similarly, perceived social goals such as being polite affect interpretations. (Pighin & Bonnefon, 2011) found that uncertainty expressions used to convey the likelihood of an undesirable medical outcome were interpreted to refer to higher event probabilities than when the same expression was used to convey the likelihood of a desirable outcome, presumably because listeners attributed the use of the uncertainty expression to sugarcoating an potentially upsetting truth.

Considering that variability in use can sometimes but not always be attributed to a specific speaker, it would be beneficial for listeners to selectively adapt to speakers only when variability cannot be attributed to other contextual factors. To what extent this is happening and to what extent contextual factors modulate listener's adaptive behavior is the focus of the work presented here.

To illustrate how contextual factors may modulate adaptive behavior, consider a speaker *S* who is in a very good mood and wants to be encouraging. If *S* tells a listener *L* "you'll *probably*

win the sweepstake" when there is only a 60% chance of winning, *L* may consider *S*'s use of *probably* instead of a weaker alternative such as *might* to be the result of *S*'s mood. Consequently, *L* would not necessarily expect *S* to use *probably* to describe the same event probability when *S* is in a worse or more discouraging mood.

Recent computational models suggest that such modulation is possible. Schuster and Degen (2020) proposed a computational model of adaptation to variable use of uncertainty expressions based on Bayesian belief updating. According to this model, when interacting with a speaker and observing their language use, listeners integrate their prior beliefs about the speaker's semantic representations and lexical preferences with the observed utterances to arrive at updated speaker-specific production expectations. Similar Bayesian belief-updating models have been proposed for adaptation in other linguistic domains, including in phonetic adaptation (Kleinschmidt & Jaeger, 2015), syntactic adaptation (Kleinschmidt, Fine, & Jaeger, 2012), and prosodic adaptation (Roettger & Franke, 2019). All of these models predict that the extent to which listeners adapt depends on how they initially expect a speaker to use language. Consequently, contextual factors that affect listeners' expectations about language use should also affect how much listeners adapt to specific speakers. If, given contextual information, a speaker's behavior matches prior expectations, there is no need to adapt.

In addition to the model-predicted influence of contextual factors on adaptation, there is empirical evidence from phonetic adaptation: Kraljic, Samuel, and Brennan (2008) used a lexical retuning paradigm to investigate adaptation to a speaker who produced a sound that was ambiguous between /s/ and /sh/. They found that without additional information listeners adapted, such that their perceptual boundary between /s/ and /sh/ shifted. However, when participants were shown a picture of the speaker with a pencil in their mouth, they explained away the observed signal as a pencil-distorted /sh/-sounding /s/ rather than as an intentionally produced /sh/-sounding /s/. Consequently, they did not adapt, i.e., their perceptual boundary between /s/ and /sh/ did not shift.

In this work, we investigate for the first time whether listeners explain away otherwise

unexpected behavior at the lexical level if they are presented with contextual information that provides a reason for a speaker's productions. Concretely, we investigate whether one contextual factor – the speaker's mood – provides such a reason for otherwise less expected uses of the uncertainty expressions *might* and *probably* (EXPLAINING AWAY HYPOTHESIS). However, considering that previous experimental studies on semantic-pragmatic adaptation (Yildirim et al., 2016; Schuster & Degen, 2020) kept all aspects of the context constant between the exposure and test phase, it could also be that listeners simply learn associations between the use of uncertainty expressions and speakers, independent of other contextual information (ASSOCIATIVE HYPOTHESIS).

We first discuss how additional contextual factors can be modeled within the computational framework proposed by Schuster and Degen (2020) and how this model may lead to "explaining away" (Pearl, 1988) of otherwise unexpected variability (Section 2). We then investigate this issue empirically as follows. We first establish that language users have different expectations about a generic speaker's use of uncertainty expressions depending on their beliefs about the speaker's mood (Exp. 1, Section 3). We then investigate how much participants adapt when they are provided with information about the speaker's mood that makes their use of uncertainty expressions more expected, and compare participants' adaptation behavior to a neutral adaptation setting in which participants do not receive any information about the speaker's mood (Exp. 2, Section 4). Finally, we investigate the relationship between adaptation and highly unexpected behavior by exposing participants to a speaker whose use of uncertainty expressions is incongruent with their mood (Exp. 3, Section 4). We find that listeners adapt less when they are presented with a reason for the speaker's behavior. However, surprisingly, we also find that listeners do not adapt more when the behavior is highly unexpected given contextual information, potentially suggesting that listeners draw additional inferences when encountering highly unexpected behavior or that there are limits on adaptation.

Modeling explaining away

In this section, we first briefly review the Bayesian belief updating model of semantic adaptation that will form the basis of the explaining away model. We then review the concept of explaining away in probabilistic graphical models, and finally we show how additional contextual factors can be integrated into the belief updating model and how the resulting model gives rise to explaining away.

Bayesian belief updating model. Schuster and Degen (2020) present a Bayesian cognitive model at the computational level (Marr, 1982; Anderson, 1990) of how listeners adapt to variable use of uncertainty expressions. The core idea of this model is that listeners have beliefs about how a generic speaker would use uncertainty expressions to describe probabilities of a future event and that in interaction, when listeners observe a specific speaker's behavior, they refine their beliefs starting off from the beliefs about the generic speaker to better match that speaker's language use. More formally, this model assumes that listeners have beliefs about speaker-specific parameters Θ_S in the form of a distribution $P(\Theta_S)$. Given observations O about the use of uncertainty expressions, listeners update their beliefs about speaker-specific parameters through Bayesian belief updating:

$$P(\Theta_S \mid O) \propto P(\Theta_S)P(O \mid \Theta_S)$$

For predicting listeners' expectations about a speaker's language use, Schuster and Degen (2020) employ a probabilistic pragmatic model within the Rational Speech Act framework (Goodman & Frank, 2016), which is parameterized by the set of speaker-specific parameters Θ_S. Θ_S is comprised of parameters guiding the semantic representation of uncertainty expressions as well as the speaker preferences for different uncertainty expressions but for the purpose of this work, we limit the discussion to an abstract set of speaker-specific parameters Θ_S (see the original paper for how individual parameters influence listeners' expectations and interpretations).

As mentioned above, this model was used by Schuster and Degen (2020) to predict participants' expectations of a speaker and their interpretations after exposure to either a

"*cautious*" or a "*confident*" speaker, whose use of uncertainty expressions differed. When talking about the probability of getting a blue gumball from a machine filled with orange and blue gumballs, the "*cautious*" speaker always used "You *might* get a blue one" to describe a probability of 60% and the "*confident*" speaker always used "You'll *probably* get a blue one" to describe a 60% probability. This exposure shifted listeners' expectations and this shift is predicted by the model as schematically illustrated in the upper part of Figure 1: Listeners start out with their beliefs about a generic speaker (center panel) and depending on the exposure speaker, their expectations about the use of *might* and *probably* either shift to cover higher probabilities ("*cautious*" speaker, left panel) or lower probabilities ("*confident*" speaker, right panel). Here and throughout this paper, we assume that an average speaker's use of uncertainty expressions is captured by the parameterization Θ_{avg}, and that the use by "*cautious*" or "*confident*" speakes are captured by the parameterizations Θ_{cau} and Θ_{con}. Thus, listeners' beliefs $P(\Theta_S)$ should assign a high probability to $P(\Theta_S = \Theta_{avg})$ prior to exposure and a high probability to $P(\Theta_S = \Theta_{cau})$ or $P(\Theta_S = \Theta_{con})$ after exposure, which is predicted by the belief updating model.

Explaining away. The second ingredient to modeling the modulation of additional contextual factors on the adaptation process is the notion of "explaining away." Explaining away is a concept from probabilistic graphical models (PGMs; Pearl, 1988), a class of models that was originally developed for uncertain reasoning in artificial intelligence systems. PGMs can be used to model causal structures between variables of interest. Figure 2 shows an example of a PGM with three variables representing whether the ground is shaking (SHAKE), whether there is an earthquake (QUAKE), or whether there is a giant roaming around (GIANT). This PGM further encodes the causal relations between the three variables: if there is an earthquake or a giant roaming around (or both), then the ground is shaking. While earthquakes are very unlikely to occur at any given point in time, it seems reasonable to assume that the sighting of a giant is still much much more unlikely, so the prior probabilities for GIANT and QUAKE intuitively follow the ordering $P(GIANT) <<< P(QUAKE) <<< 1$. According to this model, if one observes the ground shaking (and if one assumes that QUAKE and GIANT are the only possible causes for that),

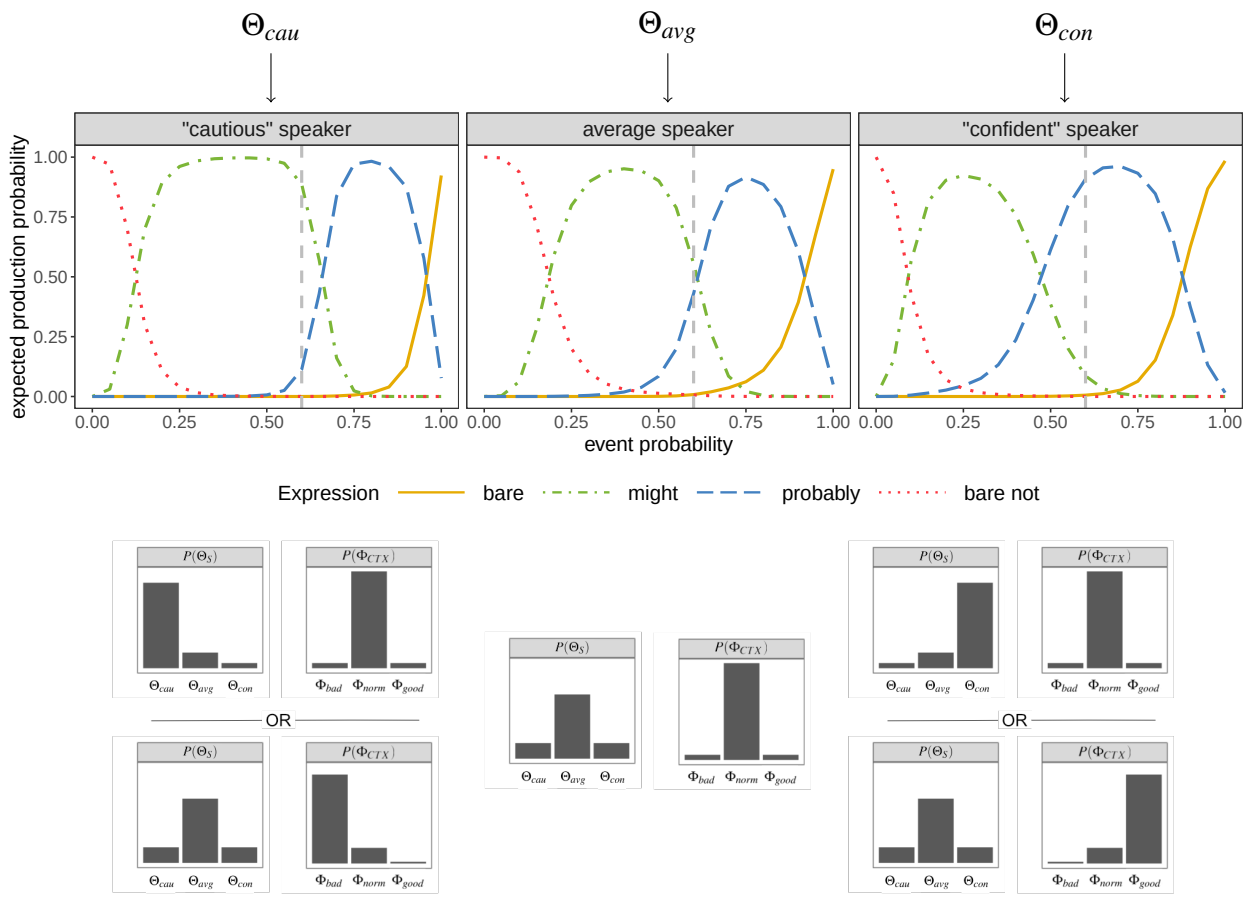

Figure 1. **Top**: Schematic visualization of different parameterizations (Θ_{cau}, Θ_{avg}, Θ_{con}) of the production expectation model. Here we assume that the four possible utterances are "You won't get a blue one" (bare not), "You might get a blue one" (might), "You'll probably get a blue one" (probably), and "You'll get a blue one" (bare). The center panel shows the expectations of an average speaker (parameterized by Θ_{avg}) that a listener may assume of a generic speaker that they haven't interacted with. The left and right panels show expectations after exposure to either a "cautious" or a "confident" speaker, parameterized by Θ_{cau} and Θ_{conn}, respectively. The vertical dashed line indicates an event probability of 60%.

Bottom: Schematic distributions over speaker-specific parameters $P(\Theta_S)$ and contextual parameters $P(\Phi_{CTX})$ that lead to expectations in the corresponding upper panels.

then either QUAKE or GIANT has to be true, and their respective probabilities will be proportional to their prior probabilities, meaning that in the absence of further information, the likelihood

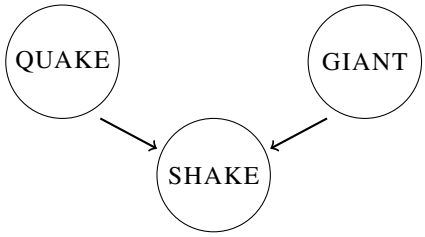

Figure 2. Example probabilistic graphical model indicating the causal structures between three variables.

$P(QUAKE)$ will be close to 1 considering that the prior probability of QUAKE is much greater than the one of GIANT. This captures the intuition that one is much more likely to attribute a shaking ground to an earthquake than a giant roaming around. However, if one again observes the ground to be shaking but also sees a giant roaming around (and thus the likelihood $P(GIANT)$ is 1), computing $P(QUAKE \mid SHAKE, GIANT)$ according to the structure of this PGM results in a very low likelihood for an earthquake, capturing the intuition that if there is already a cause for the ground shaking (the observed giant), then one is unlikely to attribute this observation to a second cause.

Apart from modeling causal reasoning in probabilistic artificial agents, explaining away has also been argued to apply to several phenomena in cognition, including visual object recognition (Murray, Kersten, Olshausen, Schrater, & Woods, 2002) and priming in word recognition experiments (Huber, 2008). Somewhat surprisingly though considering that PGMs were designed for causal reasoning, explaining away does not seem to properly capture human causal reasoning behavior in many instances (Morris & Larrick, 1995; Tenenbaum & Griffiths, 2002; Rehder & Waldmann, 2017, inter alia). It thus remains an open question how commonly explaining away predicts cognitive behavior to what extent explaining away happens in semantic-pragmatic adaptation.

Explaining away in semantic-pragmatic adaptation. As we mentioned above, according to the belief-updating model, listeners update their beliefs about speaker-specific parameters $P(\Theta_S)$ that guide their production expectations. Here, we extend this model and assume that that

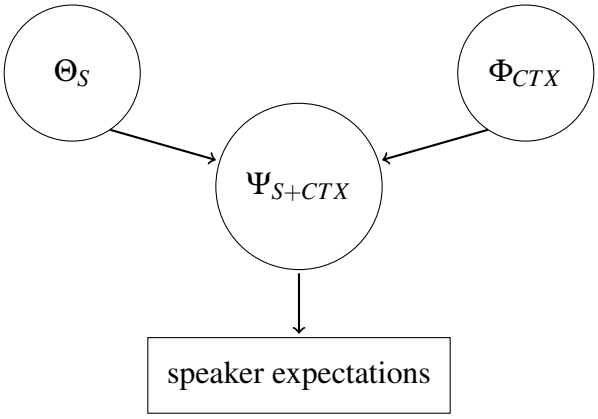

Figure 3. Probabilistic graphical model showing how speaker-specific (Θ_S) and other contextual (Φ_{CTX}) parameters are combined to form the parameters Ψ_{S+CTX} guiding expectations about a speaker's use of uncertainty expressions in a specific context.

listeners expectations are not only guided by their beliefs about speaker-specific parameters, $P(\Theta_S)$, but also by their beliefs about other contextual factors, $P(\Phi_{CTX})$. The resulting combination of factors, $P(\Psi_{S+CTX})$, is shown in the graphical model in Figure 3.[1]

To illustrate how explaining away in semantic-pragmatic adaptation may happen according to this model, consider the following simple example. Without loss of generality, we assume that there are only three possible types of speakers, an average speaker, a "*cautious*" speaker, and a "*confident*" speaker and that these three speaker types are parameterized by Θ_{avg}, Θ_{cau}, and Θ_{con}, respectively. Thus a listener's beliefs about speaker-specific parameters $P(\Theta_S)$ can be modeled with a categorical distribution over the three possible parameterizations. As for other contextual factors, we assume for the purpose of this example that only one factor, namely the speaker's mood, may influence production expectations, and that a speaker may be in a bad, neutral, or good mood, which are parameterized in the model by Φ_{bad}, Φ_{neu} , and Φ_{good}. A listener's beliefs about the speaker's mood can thus again be modeled with a categorical distribution.

[1] We leave it open how exactly beliefs about speaker-specific and other contextual parameters are combined. One possibility is that they are linearly combined like parameters in a linear model but theoretically, they could be combined through any mathematical function.

A listener's expectations about a specific speaker's use of uncertainty expressions then depends on the distributions $P(\Theta_S)$ and $P(\Phi_{CTX})$. For the purpose of this example, let us assume that the only three possible expectations that a listener may have are the ones illustrated in the upper part of Figure 1. In the lower part, we see examples of beliefs about parameterizations that lead to the corresponding expectations. For the average speaker, i.e., a generic speaker with whom a listener has no experience, we assume that listeners' beliefs $P(\Theta_S)$ and $P(\Phi_{CTX})$ are both concentrated towards the average and neutral mood parameterizations (these distributions can also be seen as listeners' prior beliefs). For the *"cautious"* speaker and *"confident"* speaker expectations, we consider two combinations of distributions each: Listeners can have expectations that differ from the average speaker because their speaker-specific beliefs $P(\Theta_S)$ deviate from the prior beliefs (upper pairs of distributions) or they can have different expectations because their beliefs about contextual factors $P(\Phi_{CTX})$ deviate from the prior beliefs (lower pairs of distributions).

How does explaining away lead to the modulation of adaptive behavior in this model? When a listener observes a speaker behaving like a "*cautious*" speaker (i.e., using *might* for a 60% event probability), there are three relevant scenarios illustrated in Figure 4: 1) The listener already has strong beliefs that the speaker's use can be characterized by the Θ_{cau} parameterization. In this case, the expectations already match the observed behavior and there is no need to adapt. 2) The listener has no specific beliefs about the speaker-specific parameter and assumes an average speaker. In this case, the expectations do not match the observed behavior and listeners will update parameters. If, as we assume in the distributions in the center of Figure 4, the prior probability $P(\Theta_{cau})$ is higher than the prior probability $P(\Phi_{bad})$, then a listener will primarily update their beliefs about speaker-specific parameters $P(\Theta_S)$ rather than the beliefs about contextual factors $P(\Phi_{CTX})$. This is the classic speaker-specific adaptation scenario in which listeners update their beliefs about $P(\Theta_S)$. 3) The listener has strong beliefs about the speaker's mood, i.e., they assign a high probability to Φ_{bad}. in this case, the expectations again match the observed behavior and the listener does not update their beliefs about $P(\Theta_S)$. Thus, importantly, if

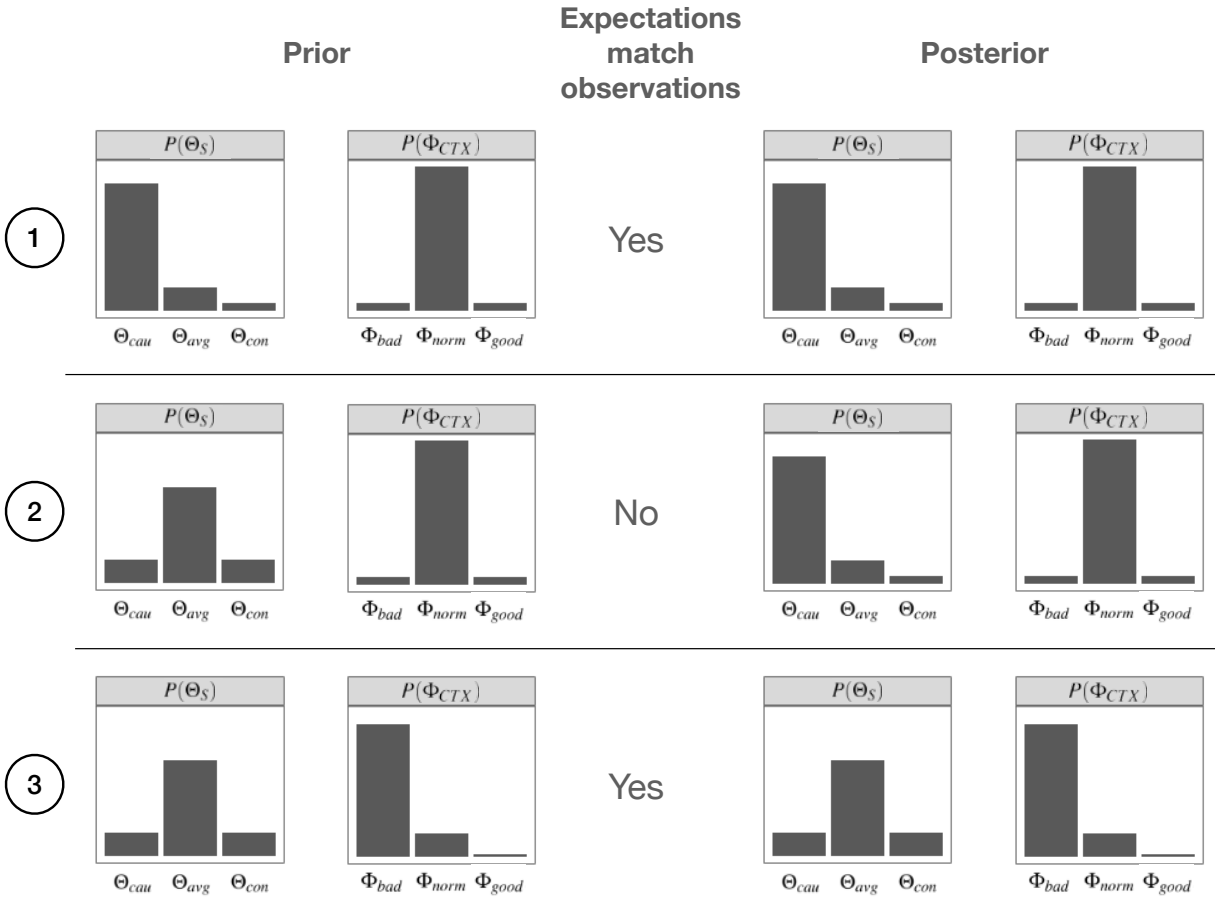

Figure 4. Three different scenarios of the adaptation process in response to a "*cautious*" speaker. In scenarios 1 and 3, the speaker's behavior matches expectations and therefore the posterior beliefs are the same as the prior beliefs. In scenario 2, the observations are unlikely under the prior expectations and therefore beliefs are updated as a result of adaptation.

the situation changes and the listener no longer has strong beliefs about the speaker's mood, the listener's expectations may revert to an average speaker. In this scenario, analogously to the ground shaking example above, the cause for the speaker's behavior is attributed to the mood and therefore the inference that this is likely a "*cautious*" speaker does not happen.

In the following sections, we test assumptions and predictions of this model.

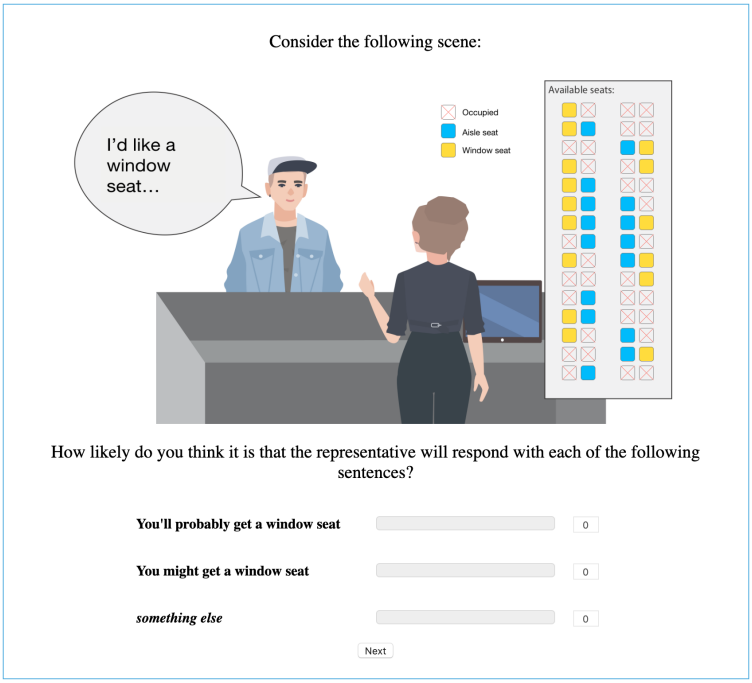

Figure 5. Example trial from Experiment 1 and the post-exposure blocks from Experiments 2 and 3.

Experiment 1: Effect of speaker mood

In Exp. 1, we investigated how one contextual factor, the speaker's mood, affects listeners' expectations about a speaker's use of the uncertainty expressions *might* and *probably* for a range of event probabilities. The choice to manipulate the speaker's mood was guided by the intuition that listeners expect a speaker in a good mood to use uncertainty expressions differently from a speaker in a bad mood. Moreover, mood is a non-inherent property of speakers that can change over time, which is important for the main research questions we are investigating here.

Procedure, materials, analyses, exclusions and predictions were pre-registered on OSF (`http://osf.io/anonymized`).

Methods

Participants. We recruited 60 participants (20 per condition) from Amazon's Mechanical Turk. We required participants to have a US-based IP address and an approval rating of at least 95%, as well as to complete a CAPTCHA at the beginning of the experiment. Participants were

paid USD 2.20 (approximately USD 12-15/hr).

Materials and procedure. At the beginning of the experiment, participants were introduced to an airline representative. Depending on the condition, the instructions explained that the representative was having a particularly bad day and feeling pessimistic and angry (*pessimist* condition); that she was having a particular great day and feeling optimistic and helpful (*optimist* condition); or that she was having a normal day (*neutral* condition). In addition to the textual mood information, the drawing of the representative showed her with an angry face (*pessimist*), a big smile (*optimist*), or a neutral facial expression (*neutral*).

Participants were then instructed that they would see scenes in which a customer of a cheap airline, who had the choice between getting a seat assigned at random or paying $50 to pick their seat, would ask the representative about their possible seat assignment, to determine the likelihood of getting their preferred seat without paying. As shown in Fig. 5, participants could see the seat map and thus determine the number of available window and aisle seats and estimate the probability of getting the preferred seat. On each trial, participants had to indicate how likely they considered the representative to respond with one of the following two utterances:

- You might get a window seat/an aisle seat. (MIGHT)
- You'll probably get a window seat/an aisle seat. (PROBABLY)

Participants indicated their production expectations by distributing 100 points across these two utterances using a slider. If they thought that neither of the two utterances were likely responses, they could assign points to a blanket *something else* option. Participants completed 36 trials: they provided 4 ratings for each of 9 different probabilities of getting a preferred seat, ranging from 0% to 100% as indicated by the seat map. Trials were counterbalanced on whether the customer asked for a window or an aisle seat and trial order was randomized.

Analysis and exclusions. Following Schuster and Degen (2020), we quantified the production expectations for MIGHT and PROBABLY by fitting a spline with 4 knots for each participant and expression and computing the area under the curve (AUC) for each of these

splines. A larger AUC indicates that an expression was rated highly for a larger range of event probabilities. To compare production expectations across conditions, we computed the difference in AUC between MIGHT and PROBABLY and compared the average difference across participants in the two conditions.

We excluded participants who provided random responses. Concretely, we excluded participants whose ratings for different event probabilities highly correlated ($r > 0.75$) with their average rating, suggesting that they always provided approximately the same rating independent of the event probability. Based on this criterion, we excluded 7 participants (*optimist*: 1, *pessimist*: 3, and *neutral*: 3).

Predictions. We predicted that participants expect the *optimistic* speaker to be encouraging and therefore to use PROBABLY for a wider range of event probabilities than the *pessimistic* speaker. Conversely, we predicted that participants expect the *pessimistic* speaker to use MIGHT for a wider range of probabilities than the *optimistic* speaker. This prediction should be reflected in larger AUC differences in the *pessimist* condition than in the *optimist* condition. Since it was unclear what mood participants attributed to the *neutral* speaker, we only predicted that the mean AUC difference for this third condition should lie between the mean AUC differences of the *optimist* and *pessimist* conditions, with the possibility of being equal to one of the two conditions.

Results and discussion

Fig. 6 shows participants' mean ratings for MIGHT and PROBABLY across the three conditions. As expected, we observe ratings close to 0 for both utterances when there is a 0% chance of getting the preferred seat (where there is a preference for the *something else* option, not shown), high ratings for MIGHT for low event probabilities, and high ratings for PROBABLY for high event probabilities. At the same time however, there are also differences across conditions. As the left panel shows, MIGHT was rated higher in the *pessimist* condition than in the *optimist* condition for a large range of event probabilities; as the right panel shows, the opposite was true

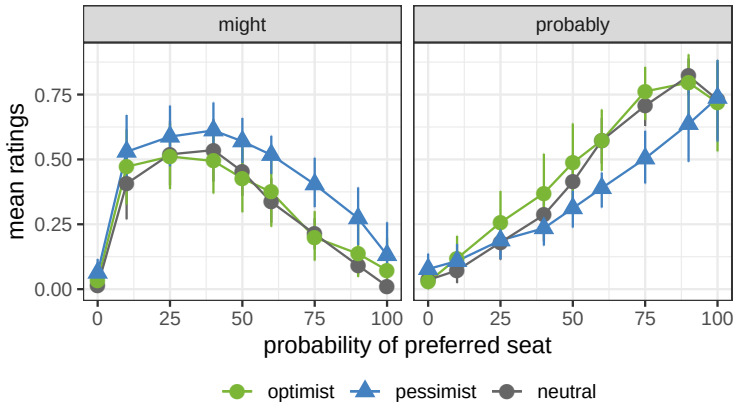

Figure 6. Mean ratings for MIGHT and PROBABLY for each condition in Exp. 1. Error bars correspond to bootstrapped 95%-confidence intervals.

for PROBABLY. Ratings in the *neutral* condition were almost identical to the ratings in the *optimist* condition. All these observations were also reflected in the AUC differences: The AUC differences in the *pessimist* condition were greater than in the *optimist* condition ($t(34) = 2.51$, $p < 0.05$). The AUC differences in the *neutral* condition – while numerically slightly larger – were not significantly different from the differences in the *optimist* condition ($t(34) = 0.35$, $p > 0.7$).

These results provide evidence that listeners have mood-dependent expectations about a generic speaker's use of uncertainty expressions. In Exp. 2 we investigate whether speaker mood affects the extent to which listeners adapt to that speaker's use of uncertainty expressions.

Experiment 2: Explaining away

In an exposure-and-test paradigm, we investigated whether adaptation to a specific speaker is modulated by knowledge about the speaker's mood. We follow (Schuster & Degen, 2020) and either exposed participants to a speaker who always used MIGHT to describe an event probability of 60% (the above mentioned "*cautious*" speaker) or a speaker who always used PROBABLY to describe an event probability of 60% (a "*confident*" speaker). Based on the results of Exp. 1, the behavior of a *cautious* speaker is more expected of a speaker who is having a bad day, and the

Condition	***cautious-pessimist***	***cautious-neutral***	***confident-neutral***	***confident-optimist***
Mood	bad	neutral	neutral	good
$p = 25\%$	–		MIGHT x5	
$p = 60\%$	MIGHT x5		PROBABLY x5	
$p = 90\%$	PROBABLY x5		–	
$p = 100\%$	BARE x3		BARE x3	

Table 1

Overview of exposure utterances in Exp. 2. p indicates the proportion of preferred available seats shown on the seat map while the speaker produced the utterance. Critical trials are highlighted in gray.

behavior of a *confident* speaker is more expected of a speaker who is having a good day. We thus hypothesized that participants' beliefs about the speaker's mood influence how much they adapt: If the speaker's behavior is mood-congruent, we expected participants' prior expectations to closely match the observed behavior and adapt less than in the neutral conditions.

Procedure, materials, analyses, exclusions and predictions were pre-registered on OSF (`http://osf.io/anonymized`).

Methods

Participants. We recruited 320 novel participants (80 per condition) from Amazon's Mechanical Turk, using the same criteria as in Exp. 1. Participants were paid USD 2.60 (approximately USD 12-15/hr).

Materials and procedure. The first block of the experiment consisted of an exposure phase with 13 trials (5 critical, 8 filler). On each trial, participants first saw a scene in which a customer asked about a specific seat and a seat map which indicated the number of available window and aisle seats (see top part of Fig. 5). To make sure participants paid attention to the seat map, they were then asked to rate how likely the customer was to get the preferred seat. They then

listened to a pre-recorded response from the airline representative. Exposure trials were identical across the *cautious-pessimist* and *cautious-neutral* conditions and identical across the *confident-optimist* and *confident-neutral* conditions but differed across these two pairs of conditions (see Table 1 for an overview): In the *cautious-pessimist* and *cautious-neutral* condition, there were 5 critical trials in which the representative described a 60% probability of getting the preferred seat with "You might get one" (MIGHT); in the *confident-optimist* and *confident-neutral* conditions, the speaker responded with "You'll probably get one" (PROBABLY). 5 filler trials in the *cautious-pessimist*/*cautious-neutral* conditions consisted of *probably* responses when there was a 90% preferred seat probability, and 5 filler trials in the *confident-optimist*/*confident-neutral* conditions combined *might* with a 25% preferred seat probability. Finally, 3 additional filler trials in all four conditions consisted of the response "You'll get one" (BARE) when it was 100% likely for the customer to get their preferred seat. Filler trials were intended to boost credibility of the speaker.

The exposure block was followed by another instruction, informing participants in all conditions that it was a week later and that the airline representative was having a normal day, followed by another manipulation check asking participants to rate how they thought the representative was feeling.

The last block of the experiment was identical to the trials in Exp. 1: participants completed 36 trials and rated how likely they thought it was that the speaker produced MIGHT, PROBABLY or *something else* for 9 different preferred seat probabilities.

Analysis and exclusions. We computed the AUC difference between the splines for MIGHT and PROBABLY for each participant as in Exp. 1. We again excluded data from participants providing random responses, which led to 52 exclusions (*cautious-pessimist*: 9, *cautious-neutral*: 14, *confident-optimist*: 18, *confident-neutral*: 11).

Predictions. We expected that participants would adapt to the use of uncertainty expressions by the different speakers and update their expectations, as illustrated in scenario 2 in Figure 4. Further, in line with the EXPLAINING AWAY HYPOTHESIS, participants in the

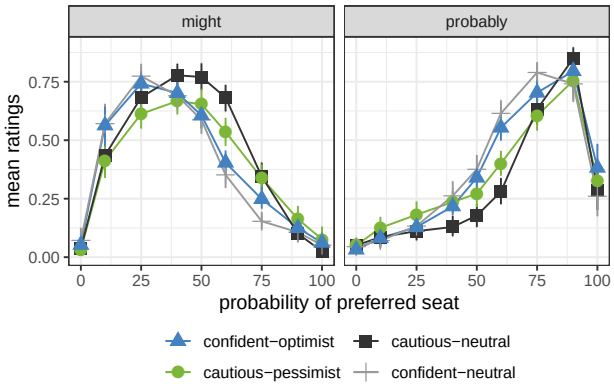

Figure 7. Mean ratings for MIGHT and PROBABLY for each condition in Exp. 2. Error bars correspond to bootstrapped 95%-confidence intervals.

cautious-pessimist and *confident-optimist* conditions, whose prior expectations should already closely match the speaker's behavior, should adapt less than participants in the other two conditions (scenario 3 in Figure 4). In terms of the AUC difference (AUC(might) - AUC(probably), we therefore expected the following ordering: *cautious-neutral* $<$ *cautious-pessimist* $\leq$ *confident-optimist* $<$ *confident-neutral*.

However, in Exp. 1, we also found that the ratings in the *neutral* condition did not significantly differ from the ratings in the *optimist* condition. This suggests that listeners' initial expectations about the speaker's mood and the associated production expectations only slightly differ across the *confident-optimist* and the *confident-neutral* conditions and therefore we also expected similar adaptation behavior in these two conditions.[2] We therefore, while expecting the numerical ordering described above, expected and pre-registered only significant differences between the *cautious-pessimist* and *cautious-neutral* conditions, and between the *cautious-neutral* and *confident-neutral* conditions.

If listeners' adaptation behavior is not affected by contextual information, in accordance

[2] This intuition was further confirmed in a pilot study with 10 participants per condition, which we conducted prior to pre-registration. In the pilot, we found the expected ordering for the *cautious-neutral*, *cautious-pessimist*, and *confident-neutral* conditions but the difference between the *confident-optimist* and *confident-neutral* condition was so small that we would have needed more than 205 participants per condition to achieve power of $\beta = 0.8$.

with the ASSOCIATIVE HYPOTHESIS, there should be no difference between the *cautious-neutral* and *cautious-pessimist* conditions and no difference between the *confident-neutral* and *confident-optimist* conditions.

Results and discussion

Fig. 7 shows the mean ratings for MIGHT and PROBABLY for the four conditions. The results are consistent with the predictions according to the EXPLAINING AWAY HYPOTHESIS: First, participants in the *confident-neutral* condition rated PROBABLY higher for a larger range of event probabilities than in the *cautious-neutral* condition and the opposite was true for MIGHT. This pattern is also reflected in the mean AUC difference, which is larger in the *cautious-neutral* condition than in the *confident-neutral* condition ($t(133) = 5.18$, $p < 0.001$). This result replicates the adaptation effect found by Schuster and Degen (2020) and suggests our seat map paradigm is suited for studying adaptation in the use of uncertainty expressions.

Second, we also observe differences between the *cautious-neutral* and *cautious-pessimist* conditions. The mean AUC difference is larger in the *cautious-neutral* condition than in the *cautious-pessimist* condition ($t(135) = 2.38$, $p < 0.02$).

Third, we also observe a numeric difference between the *confident-neutral* and *confident-optimist* conditions. Numerically, the AUC difference is larger in the *confident-optimist* condition than in the *confident-neutral* condition, but not significantly so ($t(129) = 1.61$, $p = 0.11$).

Lastly, as shown in Fig. 8, participants in the *confident-optimist* and *cautious-pessimist* condition updated their beliefs about the mood after we instructed them that the speaker was now in a normal mood, suggesting that this instruction was sufficient to update participants' beliefs about the speaker's mood. As expected, participants in the two neutral conditions did not change their beliefs about the speaker's mood.

The results from this experiment provide evidence for listeners explaining away otherwise unexpected productions if they are presented with a reason for the speaker's behavior as predicted

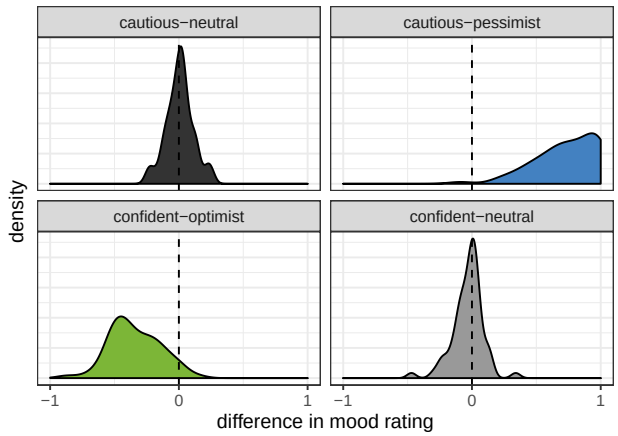

Figure 8. Differences in mood ratings in Exp. 2. The x-axis indicates the difference between the mood rating before the exposure block and the mood rating before the test block.

by the model in Section 2.

However, with additional stipulations, these results are also compatible with a simpler ASSOCIATIVE account, according to which listeners automatically track associations between speakers and language use. One aspect of the context, the speaker's mood, changed between the exposure block and the test block in the *cautious-pessimist* and *confident-optimist* conditions but not in the two neutral conditions. Therefore, it could be that this difference between the blocks leads to weaker associations between utterances and the context and therefore we observe less adaptation in the *cautious-pessimist* and *confident-optimist* conditions. To evaluate this possibility, we conducted Exp. 3.

Experiment 3: Incongruent conditions

In Exp. 3, we investigated participants' adaptation behavior when the speaker's use of uncertainty expressions was incongruent with the information about the speaker's mood, i.e., a speaker in a bad mood using uncertainty expressions like the *confident speaker* in Exp. 2, or a speaker in a good mood behaving like the *cautious speaker*. According to the EXPLAINING AWAY account, in this case listeners' prior expectations should deviate more from the observed behavior than in the neutral and congruent conditions in the previous experiment and therefore listeners

should update speaker-specific parameters more, and hence adapt more. According to an ASSOCIATIVE account, on the other hand, listeners should adapt less than in the neutral conditions because according to this account, the smaller adaptation effect that we found in the *confident-optimist* and *cautious-pessimist* conditions in the previous experiment was caused by a difference between the exposure phase and the test phase, which is still present in this experiment.

Methods

Participants. We recruited novel 160 participants (80 per condition) from Amazon's Mechanical Turk, using the same criteria as in Exp. 1. Participants were paid USD 2.60 (approximately USD 12-15/hr).

Condition	***confident-pessimist***	***cautious-optimist***
Mood	bad	good
$p = 25\%$	MIGHT x5	–
$p = 60\%$	PROBABLY x5	MIGHT x5
$p = 90\%$	–	PROBABLY x 5
$p = 100\%$	BARE x3	BARE x3

Table 2

Overview of exposure utterances in Exp. 3. p indicates the proportion of preferred available seats shown on the seat map while the speaker produced the utterance. Critical trials highlighted in gray.

Materials and procedure. The procedure was identical as in Exp. 2. There were two conditions: *cautious-optimist* and *confident-pessimist*. The *cautious-optimist* condition showed a speaker in a good mood who produced the same utterances as the *cautious-pessimist* and *cautious-neutral* speaker from the previous experiment. The *confident-pessimist* condition showed a speaker in a bad mood who produced the same utterances as the *confident-optimist* and *confident-neutral* speakers in Exp. 2. See Table 2 for an overview of the exposure trials.

Analysis and exclusions. Analyses and exclusions were identical to the ones of Exp. 2. We excluded 27 participants (*cautious-optimist*: 15, *confident-pessimist*: 12).

Predictions. We predicted that participants would adapt to different uses of uncertainty expressions: We expected the AUC difference in the *cautious-optimist* condition to be larger than in the *confident-pessimist* condition. We further predicted that listeners' prior expectations deviated more from the observed behavior than in the neutral conditions in Exp. 2. We therefore also predicted the AUC difference in the *cautious-optimist* condition to be larger than in the *cautious-neutral* condition, and the AUC difference in the *confident-pessimist* condition to be smaller than in the *confident-neutral* condition.

Results and discussion

Fig. 9 shows the mean ratings for MIGHT and PROBABLY for the two conditions in this experiment as well as the mean ratings from the neutral conditions from Exp. 2. As this plot shows, participants adapted to the different uses of uncertainty expressions. The mean AUC difference in the *cautious-optimist* condition was larger than in the *confident-pessimist* condition ($t(131) = 5.90$, $p < 0.001$). However, unexpectedly, participants did not adapt more in the incongruent conditions than in the neutral conditions. The mean AUC difference in the *cautious-optimist* condition was not larger than in *cautious-neutral* condition ($t(129) = 0.004$, $p = 0.99$), and the mean AUC difference in the *confident-pessimist* condition was not significantly smaller than in the *confident neutral* condition ($t(135) = -1.18$, $p = 0.24$).

In this experiment, we again replicated the adaptation effect. However, we did not find a significantly stronger adaptation effect across the two incongruent conditions in this experiment as compared to the neutral conditions from the previous experiment, despite the fact that listeners' expectations should have deviated more from the observed behavior.

What do these results imply for the EXPLAINING AWAY and ASSOCIATIVE accounts that we presented above? Together with the results from Exp. 2, the results from this experiment are unexpected under the ASSOCIATIVE account: if the reason for participants adapting less in the

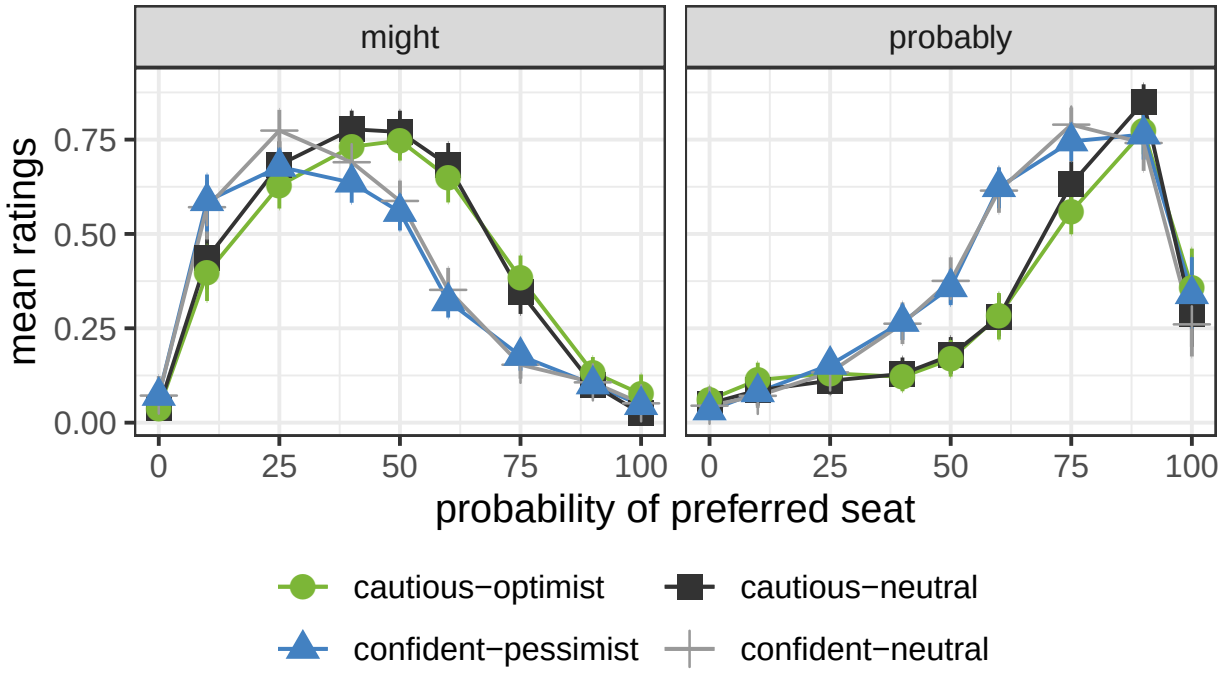

Figure 9. Mean ratings for MIGHT and PROBABLY for the two conditions in Exp. 3 as well as the neutral conditions in Exp. 2. Error bars correspond to bootstrapped 95%-confidence intervals.

cautious-pessimist condition had been the difference in context between the exposure and test blocks, we would have expected less adaptation in this experiment as well.

However, we also did not find stronger adaptation, as we would have expected under the EXPLAINING AWAY account. We can only speculate about the reasons for this, but several explanations seem likely. First, given that there was a numerical difference between the *confident-pessimist* and *confident-neutral* conditions in the expected direction, it could be that our experiment was underpowered to detect a potentially very small effect. However, a power analysis suggests we would need more than 500 participants per condition to achieve power of $\beta = 0.8$ and that this effect is too small to be detected with this paradigm with a reasonable number of participants.

Second, it could be that there is a limit on how much listeners can adapt and that this limit is already reached in the neutral conditions. If this was the case, there could still be a larger mismatch between prior expectations and observed language use when the behavior is

incongruent with contextual factors but this larger mismatch still does not lead to stronger adaptation.

Third, considering that Experiment 1 showed that participants exhibit uncertainty how a generic speaker in a good mood or a generic speaker in a bad mood uses uncertainty expressions, it could also be that listeners are adapting both to the speaker-specific language use (i.e., updating beliefs about Θ_S) and how the speaker's uses uncertainty expressions when they are in a specific mood (ie., simultaneously updating beliefs about Φ_{CTX}). Considering that the speaker's mood is different between the exposure phase and the test phase, what has been learned about the mood-dependent parameters is not relevant when making predictions during the test phase.

General Discussion

We started off this investigation with the question of how automatically adapt to variable use of uncertainty expressions: Do listeners always attribute variability to the speaker or do they incorporate contextual factors other than the speaker, resulting in a modulation of the adaptation behavior by additional contextual factors? In three experiments, we showed that language users have different expectations about the use of uncertainty expressions depending on their beliefs about the speaker's mood, and that this difference in expectations affects the extent of semantic adaptation. The results all suggest that listeners explain away otherwise unexpected behavior when they are presented with a cause, similarly as Kraljic et al. (2008) found for phonetic adaptation, and that additional contextual factors can modulate adaptive behavior at the semantic-pragmatic level.

Our results further are largely compatible with the model presented in Section 2. We confirmed the assumption that listeners expectations differ depending on contextual factors such as the mood, and we confirmed the prediction by the model that listeners update speaker-specific parameters less when prior expectations are already closely matching the observed behavior.

However, the lack of a difference between the neutral conditions and the incongruent conditions in Experiment 3 was not predicted by the model. This issue merits more investigations

and we only provided speculative reasons for the observed lack of an effect. Nevertheless, in combination with the results in Experiment 2, these results suggest that modulation of speaker-specific behavior by contextual factors happens. It is just that the exact properties of this process may be more complex than presented here and may involve additional learning or limits of adaptation.

Implications for adaptation accounts. Our experiments provide important data to evaluate existing accounts of adaptation and more generally partner-specific linguistic behavior. First, our results are not fully compatible with accounts of language processing that assume that partner-specific behavior is caused by automatic alignment of linguistic representations. Pickering and Garrod (2004), for example, argue that partner-specific linguistic behavior arises as a by-product of residual activation of linguistic representations. If we assume that listeners activate lexical representations jointly with representations of the world state, then such an account can predict that listeners adapt because after the exposure phase, there will be residual activation of representations for uncertainty expression-event probability pairs and therefore listeners' representations will be closer aligned to the speaker's representations. However, such an account does not predict the modulation of adaptation that we observed in Experiment 2 and without additional stipulations it remains unclear how automatic alignment would be inhibited by contextual factors.

While we posited the model within a Bayesian belief updating framework, our results are also mostly in line with accounts that attribute the amount of adaptation to the magnitude of an expectation violation. For example, Jaeger and Snider (2013) have shown that the magnitude of expectation violation predicts the size of syntactic priming effects, a result that is also predicted by connectionist models of syntactic learning (Chang, Dell, & Bock, 2006). However, also this account does not predict the lack of an effect in Experiment 3; considering that listeners should experience a larger expectation violation, this account predicts that listeners adapt more when observing the highly unexpected behavior.

Generalizability to other linguistic levels. The focus of this work was on the modulation of contextual factors on semantic-pragmatic adaptation. However, as we repeatedly mentioned, there seem to be a lot of parallels to the modulation of contextual factors in phonetic adaptation such as the lack of adaptation to a speaker having an object in their mouth or to a speaker being inebriated (Kraljic et al., 2008). Given these parallels and also the many parallels between the ideal adaptor framework (Kleinschmidt & Jaeger, 2015) and the adaptation model by Schuster and Degen (2020), the model presented above could be easily incorporated into the ideal adaptor framework to provide an explanation for the phenomena discussed by Kraljic et al. (2008).

Limitations. One limitation of the paradigm employed here (as in most adaptation studies to date) is that it is not interactive. Listeners take a very passive role during the exposure phase and therefore it remains an open question whether the results presented here transfer exactly to fully interactive language use.

Second, we only probed listeners' expectations about a speaker's language use while the actual task that a listener has to perform in conversation is to interpret the speaker's utterances. The main reason for this choice is that the production expectation experiment provides a more fine-grained view into the linguistic representation that listeners employ when interacting with a specific speaker. Further, Schuster and Degen (2020) demonstrated that listener's expectation directly map to interpretations. Thus, while we do not provide direct evidence that explaining away also affects interpretations, it is very likely that the results reported here also transfer to interpretations.

Conclusions. The work presented here provides evidence that listeners incorporate contextual cues other than the speaker when adapting to variable use of uncertainty expressions. This further highlights the dynamicity of the language comprehension system and its ability to integrate multiple contextual cues when interpreting utterances.

References

Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale: Lawrence Erlbaum Associates.

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, *106*(2).

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, *113*(2).

Clark, H. H. (1991). Words, the world, and their possibilities. In G. R. Lockhead & J. R. Pomerantz (Eds.), *The perception of structure: Essays in honor of wendell r. garner.* (pp. 263–277). doi: 10.1037/10101-016

Fine, A. B., & Jaeger, T. F. (2016). The role of verb repetition in cumulative structural priming in comprehension. *Journal of Experimental Psychology: Learning Memory and Cognition*, *42*(9).

Goodman, N. D., & Frank, M. C. (2016). Pragmatic language interpretation as probabilistic inference. *Trends in Cognitive Sciences*, *20*(11), 818–829. doi: 10.1016/j.tics.2016.08.005

Huber, D. E. (2008). Causality in time: Explaining away the future and the past. In N. Chater & M. Oaksford (Eds.), *The probabilistic mind: Prospects for bayesian cognitive science* (pp. 351–376). Oxford University Press. doi: 10.1093/acprof:oso/9780199216093.001.0001

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*(1).

Kamide, Y. (2012). Learning individual talkers' structural preferences. *Cognition*, *124*(1).

Kleinschmidt, D. F. (2019). Structure in talker variability: How much is there and how much can it help? *Language, Cognition and Neuroscience*, *34*(1), 43–68.

Kleinschmidt, D. F., Fine, A. B., & Jaeger, T. F. (2012). A belief-updating model of adaptation and cue combination in syntactic comprehension. In *Proc. of CogSci 2012.*

Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, *122*(2).

Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal?

Cognitive Psychology, *51*(2).

Kraljic, T., Samuel, A. G., & Brennan, S. E. (2008). First impressions and last resorts: How listeners adjust to speaker variability. *Psychological Science*, *19*(4).

Kurumada, C., Brown, M., & Tanenhaus, M. K. (2012). Pragmatic interpretation of contrastive prosody: It looks like speech adaptation. In *Proc. of CogSci 2012.*

Marr, D. (1982). *Vision*. San Francisco: W.H. Freeman and Company.

Morris, M. W., & Larrick, R. P. (1995). When one cause casts doubt on another: A normative analysis of discounting in causal attribution. *Psychological Review*, *102*(2), 331–355. doi: 10.1037/0033-295X.102.2.331

Murray, S. O., Kersten, D., Olshausen, B. A., Schrater, P., & Woods, D. L. (2002). Shape perception reduces activity in human primary visual cortex. *Proceedings of the National Academy of Sciences*, *99*(23), 15164–15169. doi: 10.1073/pnas.192579399

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, *47*(2), 204–238.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. Morgan Kaufmann. doi: 10.5555/534975

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*(2), 169–226. doi: 10.1017/S0140525X04000056

Pighin, S., & Bonnefon, J.-F. (2011). Facework and uncertain reasoning in health communication. *Patient Education and Counseling*, *85*(2), 169–172. doi: 10.1016/j.pec.2010.09.005

Rehder, B., & Waldmann, M. R. (2017). Failures of explaining away and screening off in described versus experienced causal learning scenarios. *Memory & Cognition*, *45*(2), 245–260. doi: 10.3758/s13421-016-0662-3

Roettger, T. B., & Franke, M. (2019). Evidential strength of intonational cues and rationaladaptation to (un-)reliable intonation. *Cognitive Science*, *43*, e12745.

Schöller, A., & Franke, M. (2017). Semantic values as latent parameters: Testing a fixed threshold hypothesis for cardinal readings of few & many. *Linguistics Vanguard*, *3*(1), 1–15.

doi: 10.1515/lingvan-2016-0072

Schuster, S., & Degen, J. (2020). I know what you're probably going to say: Listener adaptation to variable use of uncertainty expressions. *Cognition*, *203*. doi: 10.1016/j.cognition.2020.104285

Tenenbaum, J., & Griffiths, T. (2002). Theory-based causal inference. *Advances in Neural Information Processing Systems 15*, 43–50.

Wallsten, T. S., Fillenbaum, S., & Cox, J. A. (1986). Base rate effects on the interpretations of probability and frequency expressions. *Journal of Memory and Language*, *25*(5), 571–587. doi: 10.1016/0749-596X(86)90012-4

Yildirim, I., Degen, J., Tanenhaus, M. K., & Jaeger, T. F. (2016). Talker-specificity and adaptation in quantifier interpretation. *Journal of Memory and Language*, *87*.