Learning mechanisms in cue reweighting

Zara Harmon\(^a\), Kaori Idemaru\(^b\), Vsevolod Kapatsinski\(^\ast b\)

\(^a\) Department of Linguistics, University of Oregon, United States
\(^b\) Department of East Asian Languages and Literatures, University of Oregon, United States

1. Introduction

Phonetic categories are multidimensional (e.g., Harris, Hoffman, Liberman, Delattre, & Cooper, 1958; Lisker, 1978, 1986). The complexity of speech sound categories requires the speech perception system to rapidly integrate multiple acoustic cues—values on perceptual dimensions—in order to arrive at an overall percept (see Repp, 1982, for a review). Importantly, not all cues are equally influential because one of the perceptual dimensions contributing to a particular phonological contrast tends to be much more important than the others, serving as the primary cue dimension (Boersma, Escudero, & Hayes, 2003; Escudero, 2005; Francis & Nusbaum, 2002; Goudbeek, Cutler, & Smits, 2008; Holt & Lotto, 2006; Idemaru, Holt, & Seltman, 2012; Nearey, 1997a,b).

Cue reweighting is essential for second language learning as well as adjustments to different dialects because different languages and speech varieties may split the multidimensional space of possible sounds differently. An important aspect of cue reweighting is reallocation of attention across the dimensions defining the perceptual space (Boersma et al., 2003; Francis & Nusbaum, 2002; Idemaru & Holt, 2011), i.e., changing the extent of reliance on various perceptual dimensions in discriminating phonological contrasts. For example, for English speakers, voice onset time (VOT) is the perceptual dimension that primarily drives decisions regarding word-initial stop voicing, with F0 at the onset of the following vowel acting as a secondary cue dimension (Holt, Lotto, & Kluender, 2001; Idemaru & Holt, 2011; Kingston & Diehl, 1994; Whalen, Abramson, Lisker, & Mody, 1993).

Korean speakers, however, largely rely on F0 to distinguish English stops (e.g., Kim, Beddor, & Horrocks, 2002). Foreign-accented English learners of Korean fail to reverse their reliance on VOT when producing Korean stops (Kim & Lotto, 2002; Schertz, Lotto, Warner, & Cho, 2013). While for native American English listeners F3 is the primary cue for discriminating /r/ and /l/, Japanese speakers who are learning English rely primarily on F2 (e.g., Ingvalson, McClelland, & Holt, 2011; Yamada & Tohkura, 1991). Finally, Spanish and Southern British English speakers rely on duration for discriminating tense and lax vowels in English (e.g. heel vs. hill) whereas American English speakers rely mostly on spectral cues (Escudero, 2005; Gabay, Dick, Zevin, & Holt, 2015).

In all of these cases, native-like L2 perception requires the learner to upweight a secondary cue to a contrast and/or to down-weight the corresponding primary cue. The same feat is also required of a native speaker listening to foreign-accented speech produced by an L2
learner who has yet to reverse the cue weights in their production. Cue reweighting is illustrated in Table 1.

Reallocation of attention across perceptual dimensions is also an important part of developing adult-like perception during first language acquisition. In a series of studies, Nittrouer and her colleagues have demonstrated that children weight some acoustic cues differently when compared to adults. For example, in categorizing fricatives by place of articulation (e.g. sore vs. shore), 5- to 7-year-old children tend to rely more on vowel formant transition cues while adults use the peak of the frequency spectrum of fricative noise as the main cue (Nittrouer & Miller, 1997; Nittrouer, 1992, 2002). Overall, a strong body of evidence suggests that phonetic cues weight change throughout development, a process referred to as Developmental Weighting Shift (Nittrouer & Miller, 1997; see also Hazan & Barrett, 2000; Mayo & Turk, 2004; Morrongiello, Robson, Best, & Clifton, 1984; Parnell & Amerman, 1978). In general, learners often rely on the most salient or ubiquitous cues at first, but—as they learn the cue–outcome contingencies of the environment—eventually come to rely on the cues that are most informative (Bates & MacWhinney, 1989; MacWhinney, 1987, 2005).

Speech perception therefore involves constant reallocation of attention across perceptual dimensions, so that some cues to phonetic categorization are upweighted and others are downweighted. Moreover, appropriate training can lead to reallocation of attention for both second language learners (e.g., Lively, Pisoni, & Logan, 1992) and adult native speakers of English (e.g., Francis, Baldwin, & Nusbaum, 2000). Yet, the learning mechanisms responsible for these adjustments are not well understood. The present paper aims to contribute to our understanding of these mechanisms and their role in cue weighting.

### 1.1. Error-driven learning mechanisms

We focus particularly on two error-driven learning mechanisms, supervised learning and reinforcement learning. We first introduce the intuitions behind these approaches and our experimental design, and discuss the role of distributional information in learning, before proceeding to a more detailed description of the computational implementations of these mechanisms, and the experiment testing their predictions.

#### 1.1.1. Supervised learning: The Rescorla–Wagner model

The main tenet of supervised error-driven learning is that learning occurs as a result of prediction error. In other words, learning occurs when environmental events do not meet the learner’s expectations. With respect to phonetic category learning, one may learn which acoustic differences do and do not matter on the basis of feedback. One source of feedback comes from word recognition (Francis et al., 2000; Norris, McQueen, & Cutler, 2003). Suppose a listener perceives a speaker’s production as rice but then realizes that the speaker was instead referring to lice. This prediction error can give the listener information that she is using the acoustic cues in the signal incorrectly, so that some cues previously thought to predict /ɹ/ are re-interpreted as cues to /l/. Francis et al. (2000) have shown that explicit feedback can lead native speakers of English to change their reliance on either formant transitions or the burst in identifying the phonetic category of a stop (see also Hayes-Harb, 2007). Gabay et al. (2015) show that learners can also acquire novel auditory categories without explicit feedback as long as category identity is predictive of the response they need to produce. Similarly, Thiessen (2011) has shown that sound category learning is facilitated when the category is predictive of upcoming speech. Implicit lexical feedback has also been shown to alter native sound representations (Norris et al., 2003). However, even though there is considerable evidence that prediction error can help language learners weight and reweight phonetic cues, existing data do not allow us to determine exactly how prediction error is utilized in this process, nor do they inform the choice among alternative models of error-driven learning.

The most influential supervised error-driven learning model, often called the “standard theory” of associative learning, was developed by Rescorla and Wagner (1972). Though developed independently, the Rescorla–Wagner model is closely related to Widrow and Hoff (1960) delta rule that continues to be widely deployed in artificial neural networks (see Danks, 2003; Gluck & Bower, 1988; Miller, Barnet, & Grahame, 1995; Sutton & Barto, 1998). Rescorla and Wagner proposed that learning consists of acquiring a network of cue–outcome mappings that ultimately allows the learner to accurately predict outcomes from presented cues. When an outcome occurs in the presence of some cues, the connections from the present cues to the outcome are incremented by a certain amount. The increment is proportional to prediction error, i.e. how surprising the outcome’s occurrence was. If the outcome is already expected to occur when the cues are perceived, very little learning occurs. Conversely, if an outcome is unexpectedly absent after some cues are perceived, the weights of the connections from these cues to the absent outcome are decremented. The decrement is proportional to prediction error, i.e. how unexpected the outcome’s absence was.

Despite its advanced age, the Rescorla–Wagner model is enjoying a renaissance in language acquisition where it has been employed to capture a wide range of findings (e.g., Arnold, Tomaschek, Sering, Lopez, & Baayen, 2017; Baayen et al., 2011, 2016; Ellis, 2006; Olejarczuk, Kapatsinski, & Baayen, 2018; Ramscar et al., 2010, 2014; Ramscar, Dye, & Klein, 2013a). For example, Olejarczuk et al. (2018) show that the Rescorla–Wagner model can acquire regions of perceptual equivalence in sound categorization by mapping acoustic features to phonological categories. Arnold et al. (2017) show that the Rescorla–Wagner model can be used to acquire useful relationships between acoustic features and lexical items, resulting in humanlike spoken word recognition accuracy. However, the predictions of the Rescorla–Wagner model have not been directly contrasted with the predictions of alternative learning mechanisms in this literature.
some policies that are subject to selection by consequences include the allocation of attention to certain cues, dimensions, or stimuli over others (see also Roelfsema, van Ooyen, & Watanabe, 2010). As discussed above, phonetic cue weighting is also widely seen as attention allocation (Francis & Nusbaum, 2002; Francis et al., 2000; Holt, Tierney, Guerra, Laffere, & Dick, 2018). Focusing on a particular perceptual dimension can then be punished or reinforced. In fact, overt reinforcement has been shown to be efficacious in changing attention allocation across subliminal visual cues (Seitz & Watanabe, 2003; Seitz, Kim, & Watanabe, 2009).

In language learning, attending to the right cues does not often bring overt and immediate rewards. However, successful predictions can be reinforcing and prediction error can be considered a type of punishment. Given this assumption, reinforcement learning predicts that attention will be reallocated across perceptual dimensions like VOT and F0 to the extent that doing so results in a reduction in error. Importantly, because adapting to a new accent requires abandoning a policy of attention allocation that has previously worked well, reinforcement learning predicts that this can happen only if another policy is estimated to be superior. This means that attention could be reallocated from VOT to F0 only if attending to F0 is observed to reduce prediction error compared to attending to VOT.

1.1.3. Distinguishing supervised and reinforcement learning

All previous studies on error-driven reweighting of phonetic cues have been designed so that reallocating attention in the direction suggested by feedback would reduce error (e.g., Francis et al., 2000; Gabay et al., 2015; Hayes-Harb, 2007). This is a problem if we wish to distinguish between supervised and reinforcement learning mechanisms, both of which can account for learning under these conditions. An example of this kind of design is illustrated in Table 2a.

In the present experiment, we aim to discriminate between supervised and reinforcement learning by testing whether—as predicted by reinforcement learning—cue reweighting occurs only when it is observed by the learner to reduce error and is not expected when predictions result in prediction error but changes in behavior fail to reduce it. We do this by examining the efficacy of training illustrated in Table 2b.

In both Table 2a and b, D1 is no longer predictive of the correct response. However, in the training condition shown in Table 2a, D2 is predictive of the correct response. Reallocating attention from D1 to D2 will therefore reduce prediction error. In the training depicted in Table 2b, D2 is not predictive of the correct response because it is held constant at the a priori ambiguous value. Reallocating attention to D2 will therefore not reduce prediction error in this condition. Supervised learning models like Rescorla and Wagner (1972) predict down-weighting of D1 in both conditions. In contrast, reinforcement learning predicts that D1 will be downweighted only when D2 is informative (Table 2a). Constant D2 training therefore allows us to discriminate between supervised and reinforcement learning. Previous work on error-driven reweighting of phonetic dimensions (e.g., Francis et al., 2000; Gabay et al., 2015; Hayes-Harb, 2007) explored only Informative D2 training (Table 2a) and not Constant D2 training (Table 2b).

1.2. Distributional learning

Maye and Gerken (2000) proposed an additional mechanism for sound category learning, which they call distributional learning. Distributional learning refers to learning category structure from frequency distributions in a multidimensional acoustic space (see also Escudero, 2005; Maye, Werker, & Gerken, 2002; Maye, Weiss, & Aslin, 2008; Toscano & McMurray, 2010). In particular, if there are two sound categories, X and Y, in the learner’s ambient language, there should be a bimodal frequency distribution of sounds along some critical acoustic dimension distinguishing the two categories.

The difference between unimodal and bimodal distributions is thought to be particularly important for bootstrapping the acquisition of novel sound categories (Escudero, 2005; Maye et al., 2002, 2008; Toscano & McMurray, 2010). Consequently, the focus of most previous studies of distributional learning has been on the acquisition of novel phonetic contrasts, i.e., on splitting an uncategorized perceptual dimension or a familiar category in two. Because of this focus on learning to discriminate, most of the research on distributional learning has investigated the effects of exposure to a bimodal distribution.

In the present study, we investigate whether exposure to a unimodal distribution along an acoustic dimension would lead to down-weighting that dimension. The studies that have investigated the effects of exposure to a unimodal distribution have not found a significant difference in sound discrimination between that condition and a no-training control (Baese-Berk, 2010; Gureckis & Goldstone, 2008; Hayes-Harb, 2007; Maye et al., 2008). However, the lack of a significant effect of exposure to a unimodal distribution in these studies may be due to their focus on splitting a dimension. Because of this focus, participants in these studies were presented with cues that they had little reason to discriminate prior to training. This may have left little room for unimodal training to reduce discrimination further.

The learning mechanisms behind distributional learning remain a subject of controversy (e.g., Escudero, 2005; Feldman, Griffiths, & Morgan, 2009; Kleinschmidt & Jaeger, 2015; Lim, Fiez, & Holt, 2014; Olejarczuk et al., 2018; Van Rooij, Boersma, & Benders, 2015). Here, we focus on the use of distributional information in generative models of sound categorization. A generative model infers characteristics of the speaker’s production system from the experienced distributions, where a unimodal distribution suggests a single production target (e.g., Feldman et al., 2009; Kleinschmidt & Jaeger, 2015). If a speaker has only one production target along a dimension, she has little control over that dimension. Therefore, the listener can use a unimodal distribution along D1 to infer that the speaker is not using D1 to cue distinct outcomes. Thus, the training depicted in Table 3a provides the learner with distributional evidence for downweighting D1, whereas the training in Table 3b does not.

Table 3
An extreme manipulation of the distribution along the primary cue dimension (D1).

<table>
<thead>
<tr>
<th>a. Unimodal D1 training</th>
<th>b. Bimodal D1 training</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Diagram showing unimodal and bimodal distributions for D1 and D2" /></td>
<td></td>
</tr>
</tbody>
</table>

Note. Total number of training trials is equated across the two conditions. Graded values indicate cue–outcome mappings that are consistent with the learner’s prior beliefs (Table 1a).
A second reason to manipulate modality of the distribution along the primary cue dimension is that supervised error-driven learning predicts a modality effect in the opposite direction. A bimodal distribution along D1 means that the learner experiences many trials featuring cues for which she confidently makes wrong predictions. In contrast, a unimodal distribution means experiencing many trials featuring an ambiguous value of D1. When D1 is ambiguous, the learner’s native language experience leads her to expect that the outcome will be based on D2. This expectation is only confirmed by the training, resulting in relatively little prediction error. Therefore, when compared to a bimodal distribution, a unimodal distribution along D1 provides less information regarding the fact that D1 is no longer informative. A bimodal distribution along D1, as in Table 4, is therefore expected to result in a more rapid downweighting of D1 compared to a unimodal distribution (Table 3a), at least when D2 is distributed bimodally and therefore appears to be under the control of the speaker.

1.3. The present study

The present experiment combines the two manipulations illustrated in Tables 2 and 3. The specific contrast we examine is word-initial stop voicing in the words bear, pear, beer and pier. For this contrast, VOT serves as the primary cue dimension (D1) for English speakers, while F0 at the onset of the following vowel is a secondary cue dimension (D2). The primacy of VOT has been established by previous work. First, the F0 distributions between bear and pear overlap more than VOT distributions (Abramson & Lisker, 1985; Clayards, 2018; Lisker & Abramson, 1964). Second, participants exposed to a negative VOT/F0 correlation—a reversal of the English contingency—learn to downweight F0 rather than learning to downweight VOT (Idemaru & Holt, 2011). In the present study, we hoped to instead induce an upweighting of F0 at the expense of VOT and in doing so test the predictions of alternative learning mechanisms.

The experimental conditions are shown in Table 4. Throughout the paper, when referring to an individual condition, the names in the six panels of Table 4 will be used (e.g., condition Auni), but a condition will be identified only by the letter when information about modality is irrelevant (e.g., condition A). In A, which functions as the control condition, VOT was informative regarding voicing, matching the expectations of an English speaker. In contrast, F0 was held constant at an uninformative value. In B and C, VOT was no longer informative because the feedback was /b/ or /p/ 50% of the time at every value of VOT. In condition B, F0 was constant throughout training, held at the same value as in condition A. In condition C, F0 varied between two values, which were equidistant from the constant value used in conditions A and B. The feedback was always /p/ when F0 was high and /b/ when it was low. The VOT distribution manipulation is represented by bolding and a larger font, which represents high exposure frequency (see Fig. 3 for the actual unimodal and bimodal frequency distributions along VOT).

There are two ways in which this design complicates the simple designs in Tables 2 and 3. First, as shown in Table 4, the modality manipulation we used is not as extreme as that in Table 3 because such an extreme manipulation would not allow us to differentiate between conditions Auni and Buni which would both be reduced to the central cell: VOT25msF0215Hz→b|p. Second, the overall amount of prediction error was equalized between conditions A and B by randomly pairing vowel formants with vowel feedback in condition A, but not in condition B. Thus, feedback randomly indicated that b25ms ear is either bear and pear in condition B and either bear or beer in condition A, resulting in 50% prediction error. However, this prediction error is attributable to incorrectly predicting voicing from VOT in B but vowel from formants in A. As we show next, supervised learning is capable of making such blame assignments, downweighting VOT in condition B and vowel formants in condition A. On the other hand, these conditions are equivalent with respect to reinforcement: any attention allocation policy would produce 50% error in both conditions.

2. Computational simulations

We compared the two alternative approaches to error-driven learning, supervised and reinforcement learning, using computational simulation. In the interests of full disclosure, we should note that the supervised model was devised during experiment design, whereas the reinforcement learning model was devised after the experiment was run. We believed that the supervised learning model’s prediction of downweighting VOT in condition B was likely incorrect, and considered it likely that learning would be obtained only in condition C, where the

<table>
<thead>
<tr>
<th>Condition Auni</th>
<th>Condition Buni</th>
<th>Condition Cuni</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>F0</strong></td>
<td><strong>VOT</strong></td>
<td><strong>F0</strong></td>
</tr>
<tr>
<td>250</td>
<td>b</td>
<td>250</td>
</tr>
<tr>
<td>215</td>
<td>p</td>
<td>215</td>
</tr>
<tr>
<td>180</td>
<td>b</td>
<td>180</td>
</tr>
</tbody>
</table>

Note. Grayed-out symbols indicate cue-outcome mappings consistent with English, i.e. the learners’ prior beliefs. Bolding and larger font indicate high frequency during training (e.g., b|p for VOT 25 and F0 215 in Auni condition). The frequent stimuli in the three unimodal conditions carry little information compared to the frequent stimuli in bimodal conditions.
learners can observe an alternative attention allocation policy to be superior. However, we did not have a formal model instantiating these predictions at the time of experiment design.

2.1. Supervised learning

We derived the predictions of the Rescorla–Wagner model for this design by means of a computational simulation. We used VOT and F0 values as cues to voicing. For example, a stimulus with VOT of 25 ms and F0 of 215 Hz paired with feedback indicating that the intended referent was bear would be represented as 25 ms + 215 Hz + [ɛ] → /b/ + /ɛ/ (see Olejarczuk et al., 2018, for a similar use of the Rescorla–Wagner model).

We initialized the model with beliefs about the VOT/voicing relationship characteristic of English speakers. We then updated these beliefs (cue weights) given the experimental subjects’ experience using the Rescorla–Wagner rule in (1)–(2), where (1) applies to weights of connections between the present cues (C) and an outcome (O) occurring on a certain trial (at time t), and (2) applies to connections to outcomes that did not occur.

The subtraction terms in parentheses represent the prediction error: how unexpected the presence or absence of each outcome was, given the cues that were observed, i.e., the difference between the outcome’s activation by the cues and either 1 (maximum expectedness, which is the correct activation level for outcomes that occur) or 0 (minimum expectedness, the right activation level for outcomes that do not occur). Λ is the learning rate, which was set to 0.1 for this simulation. As learning rate increases, the difference between condition A on the one hand and conditions B and C on the other increases in magnitude but the ordinal pattern of differences among conditions remains unchanged.

\[
\begin{align*}
   w_{c\rightarrow o}^{t+1} &= w_{c\rightarrow o}^{t} + (1 - \sum w_{c\rightarrow o}^{t}) \times \Lambda \\
   w_{c\rightarrow o}^{t+1} &= w_{c\rightarrow o}^{t} + (0 - \sum w_{c\rightarrow o}^{t}) \times \Lambda
\end{align*}
\]

For each test stimulus, activation of an outcome like /p/ or /b/ given some cues (a value of VOT, a value of F0 and the vowel, which activates both equally) is the sum of the corresponding cue weights (\(\sum w_{c\rightarrow o}^{t}\)). The choice probabilities are then derived from the outcome activations using the Luce choice rule (Luce, 1959). First, activations below 0 are set to 0: the probability of selecting an inhibited outcome is assumed to be zero and activations above 1 are set to 1. Then, the probability of selecting /p/ is the activation of /p/ divided by the sum of activations of /p/ and /b/.

As shown in Fig. 1, the model predicted that both B and C conditions should lead participants to downweight the experienced VOT values in proportion to the frequency with which these values are experienced during training: response proportions in conditions Buni, Bbi, Cuni, and Cbi in Fig. 1 are closer to 50% than the values before training or those in the two A conditions. VOT is expected to be downweighted in conditions B and C equally, resulting in a difference between these two conditions and A. A lower weight of VOT is expected in bimodal VOT conditions. This is because the bimodal distribution along VOT exposes VOT-reliant participants to more prediction error.

The expected downweighting of VOT in condition B is a particularly interesting and counterintuitive prediction of the Rescorla–Wagner model because downweighting VOT would not improve performance in that condition. After downweighting VOT in condition B, the Rescorla–Wagner model makes 0.5 prediction error on every trial instead of making no error on half the trials and an error of 1 on the other half. From a participant’s perspective, 50% of responses would also be wrong and result in an error message both before and after the downweighting.

In addition, the Rescorla–Wagner model predicted an upweighting of VOT. Unlike downweighting of VOT, upweighting of F0 is expected to occur only in condition C. Because F0 is held constant in condition B, it is not predictive of feedback. In fact, the model learns nothing about the test values of F0 in conditions A and B, since these values are absent from training. In contrast, in condition C, F0 values presented at test are predictive of feedback throughout training. Interestingly, the increase in F0 weight in the C condition is reliably larger when the distribution of VOT values is bimodal, as is the case for downweighting of VOT.

2.2. Reinforcement learning

Our reinforcement learning model is perhaps the simplest possible implementation of the idea that learning consists of updating values of attention allocation policies (various weightings of VOT and F0) on the basis of prediction error generated by predicting phonological categories from a weighted sum of VOT and F0 values. The weights of F0 and VOT considered by the model ranged from 0 to twice the weight we considered most plausible for English based on control condition (condition A) data. That is, the learner considers the possibility of weighting the two dimensions more or less, but does not consider the possibility of reversing the relationships between VOT or F0 on the one hand and voicing on the other. This restriction implements the idea that the alternative policies are policies of attention allocation.

Before training, the most plausible policy for English was assigned a high weight (0.9), while others were assigned a low weight (0.01). The high value assigned initially to the English-like weights ensures that the policy based on the participants’ prior experience would be retained unless some other policy reduces prediction error during training.

Like the human learners, the model encountered stimuli with various VOT and F0 values. The model then predicted either /b/ or /p/ from each possible policy for the values of VOT and F0 just encountered. If the feedback provided on that trial was consistent with the prediction of a policy, the value of that policy was increased by the learning rate. Otherwise, it was decreased by the learning rate. The learning rate (0.1) was sufficiently fast for the initial, pre-training policy values not to make a difference to after-training behavior if some policy would reduce prediction error over the best pre-training policy (as is the case in condition C). At test, the model averaged the predictions of the policies that had the highest (maximum) values at the end of training. Because the predictions are limited to 0 and 1, these included a wide range of high F0 weights.

The predictions of the reinforcement learning model are shown in Fig. 2. As the figure shows, the model predicts identical performance in conditions A and B, and significant downweighting of VOT accompanied by upweighting of F0 in Condition C. This is because feedback

---

1 Setting negative activations to zero is necessary to avoid negative probabilities. If negative activations are retained, the model still predicts downweighting of VOT in condition B. In fact, it predicts that downweighting of VOT will be greater in condition B than in condition C. Setting high activations to 1 is unnecessary, but does not change the predictions of the model.

2 We ran 1000 simulations with different trial orders. The curves in Fig. 1 represent mean values but the ordinal differences between the conditions are unchanged across trial orders.
during training disfavors all attention allocation policies equally in conditions A and B, preserving the difference in values between policies established prior to training based on native language experience. In condition C, policies that reallocate attention to F0 are superior to the pre-training policy based on feedback because ignoring VOT and attending to F0 allows for almost perfect prediction. The learning rate is sufficient for this feedback to overcome the pre-training commitment to VOT.

The modality of the VOT distribution is predicted to have no effect. However, this prediction is made by the model only if value is maximized; that is, if only policies of maximum value are followed at test. Otherwise, the model agrees with the Rescorla–Wagner model in predicting faster learning in the Cbi condition compared to the Cuni condition. However, regardless of modality, no learning is expected in the B conditions. This distinguishes the reinforcement learning from the Rescorla–Wagner model regardless of parameter settings. This difference in predictions directly follows from the central claim of reinforcement learning (Sutton & Barto, 1998): changes in behavior occur only if they reduce expected future error. In Condition B, the error rate will remain the same regardless of the learner’s behavior, leading them to continue following the policy that has been most rewarding in their past (native language) experience.

To summarize, the ability to adjust cue weights is an important property of speech perception for all groups of language learners. The main contribution of this paper is to empirically compare the predictions of alternative error-driven learning mechanisms most likely involved in this process—supervised learning and reinforcement learning—and to examine the interaction between prediction error and distributional information.

3. Methods

3.1. Participants

A hundred and eighty adult native speakers of American English participated in the experiment. Participants were all undergraduate students at the University of Oregon, recruited from the Linguistics and Psychology departments’ human subject pool. To prevent selection bias, participants signed up for the experiment blindly and were assigned to one of the six conditions in the order they signed up for the experiment. Participants received course credit for participating in the experiment.

3.2. Stimuli

Natural utterances of beer, pier, bear and pear were digitally recorded (22.05 kHz) in a sound-attenuated booth by an adult female
native English speaker. The pairs of end-points were selected for similar duration (385 ms) and F0 contour. Stimuli were then constructed using progressive cross-splicing of the end-point tokens (McMurray & Aslin, 2005). A total of 18 splice points were identified in each of the voiced and voiceless end-point tokens, with approximately 2 or 3 ms increments and always at zero crossing. To create a VOT continuum step, part of the voice token was removed starting at the beginning and ending at the current splicing point. Corresponding interval from the voiceless token was extracted and inserted into the voiced token. From the resulting sounds, those with VOT values of 5, 15, 20, 25, 30, 35, and 45 ms were retained as stimuli.

The F0 of the two VOT continua (beerpier and bear-pear) were then manipulated such that the onset frequency of the vowel varied in three 35-Hz steps, 180 Hz, 215 Hz, and 250 Hz. In each sound, the F0 contour of the original production was manipulated using Praat 5.3 (Boersma & Weenink, 2010) to adjust to the target onset F0 values. From the onset, the F0 decreased quadratically to 150 Hz across the word. The high (250 Hz) and low (180 Hz) values of F0 and the contour modeled the natural production of the speaker. The stimuli were then normalized to the same root-mean-square amplitude (75 dB).

3.3. Procedure

3.3.1. Training task
Each trial began with the simultaneous presentation of the auditory stimulus and four images corresponding to the four test words (beer, pier, bear, and pear) on the computer screen. The four images appeared in the same locations in all trials: beer appeared on the top left, pier on the top right, bear on the bottom left, and pear on the bottom right of the screen. For each trial, the participant’s task was to click on the image of the word that they heard. In the training block, the response was immediately followed by the feedback. When the response was correct, the selected image remained on the screen for 1.5 s while others disappeared. When the response was incorrect, the word “Incorrect!” in red appeared while the correct image remained for 1.5 s.

3.3.2. Training
As described above, the experiment involved 6 training conditions that differed in the modality of VOT and the informativeness of both VOT and F0. These conditions were each presented with either unimodal or bimodal distribution of VOT spanning the English voiced and voiceless VOT values (Fig. 3). In what follows, we provide a detailed description of each condition.

3.3.2.1. Modality of VOT. In the three unimodal conditions (Auni, Buni, and Cuni), participants were exposed to the unimodal distribution of VOT, as shown in Fig. 3. The center value of the VOT (25 ms) was presented to the participants most frequently (20 times for beer-pier and 20 times for bear-pear stimuli), whereas VOT values of 15, 20, 30, and 35 ms were presented 16 times each for beer-pier and 16 times each for bear-pear, and VOT values of 5 and 45 ms were presented 6 times each for beer-pier and 6 times each for bear-pear in the training block. In conditions A and B, all training stimuli had a constant F0 value, 215 Hz, which is midway between the F0 values presented at test.

In the three bimodal conditions (Abi, Bbi, and Cbi), participants were exposed to the bimodal distribution of VOT. The VOT values of 5, 15, 35, and 45 ms were presented to the participants most frequently (16 times each for beer-pier and 16 times each for bear-pear stimuli), whereas VOT values of 20 and 30 ms were presented 12 times each for beer-pier and 12 times each for bear-pear and VOT value of 25 ms was presented 8 times for beer-pier and 8 times for bear-pear in the training block. Note that the frequencies of 15 ms and 35 ms VOTs (a typical voiced value and a typical voiceless value in English) were identical during training (32 times for both beer-pier and bear-pear stimuli) in both conditions. These stimuli were therefore used in the test block to assess the effect of distribution independently of the effect of frequencies of the specific VOT values presented (see also Maye et al., 2002, 2008). The maximally ambiguous 25 ms VOT value was also presented at test.

3.3.2.2. Feedback: Informativeness of VOT and F0. Condition A: Informative-VOT Constant-F0. In this condition, the onset F0 of the voiced was held constant at 215 Hz so that feedback provided no information regarding whether variation in F0 co-varies with voicing. In contrast, VOT was clearly informative about voicing according to information regarding whether variation in F0 co-varies with voicing. Thus, this condition was added as a control condition for condition B because both of these conditions introduced random mapping between the vowels in the stimuli ([i] vs. [e]) and vowel category (beer-pier vs. bear-pear) had a random relationship. In other words, when participants heard a beer-pier stimulus with 15 ms VOT, the correct response would be beer half the time and bear half the time. Responding with either pier or pear would always be wrong. When participants heard the bear-pear stimulus with 35 ms VOT, the correct response would be pear half the times and pier half the times. Responding with beer or bear would always be wrong.

This condition was added as a control condition for condition B because both of these conditions introduced random mapping between an acoustic dimension and a phonetic category. In condition B, this random mapping was between VOT and voicing while in condition A, it was between vowel formants and vowel categories. However, in condition A, as in English, VOT is robustly predictive of voicing. Thus, categorization of the test stimuli in this condition should reflect the perceptual weight of VOT when VOT is predictive of voicing.

Condition B: Uninformative-VOT Constant-F0. In this condition, the stimuli still had a constant onset F0 of 215 Hz, providing no information regarding the informativeness of variation in F0. However, feedback was designed to convey to the learners that VOT is no longer predictive of voicing or lexical identity. For each VOT value, the correct response was set as /b/-words (beer and bear) half the time and /p/-words (pier and pear) half the time. Unlike in the control condition A, vowel feedback was always consistent with English, thus responding bear to beer or pear to pier was always incorrect. This condition is crucial because supervised learning predicts that the outcome of learning in this condition should be different from the A condition, while reinforcement learning predicts that it should not.

---

![Fig. 3. Bimodal (solid line) and unimodal (dashed line) distributions of VOT during training.](image-url)
Condition C: Uninformative-VOT Informative-F0. In this condition, each VOT value was still mapped onto /b/ and /p/ 50% of the time. The unique feature of this condition was that the training stimuli varied across two levels of F0: 180 Hz (low F0) and 250 Hz (high F0). In order to maintain the random VOT–voicing mapping, F0 and VOT were decorrelated: each value of VOT was mapped half the time onto F0 of 180 Hz and half the time onto F0 of 250 Hz. For example, the beer-pier stimulus with 25 ms VOT and low F0 was presented 10 times and the beer-pier stimulus with 25 ms VOT and high F0 was presented 10 times. F0 was perfectly predictive of voicing during training: for a stimulus, the /b/ response was correct when F0 was low and the /p/ response was correct when it was high, regardless of the VOT value. These F0–voicing mappings are the same as in English but, because the VOT–voicing relationship is random, F0 is the primary cue to voicing in this training condition. Both supervised and reinforcement learning predict that this condition should produce more downweighting of VOT than condition A. In addition, reinforcement learning predicts more downweighting of VOT than in condition B. Distributional learning predicts more downweighting of VOT in Cuni than in Cbi, whereas supervised learning predicts the opposite. All learning mechanisms predict that F0 should be upweighted in this condition.

3.3.3. Test
Following each training block (192 trials), participants completed one block of word identification without feedback. VOT values of 15, 25, and 35 ms were used during test. Of particular importance, however, were stimuli with the VOT values of 15 and 35 ms, which were presented to participants an equal number of times in training regardless of the modality of VOT. If participants respond to stimuli with these VOT values differently depending on the distribution of VOT, the observed differences could be attributed to the modality of VOT—a prediction of distributional learning (Maye et al., 2002). The test stimuli were presented in random order for a total of 60 test trials (3 VOTs × 2 F0s × 2 words × 5 repetitions = 60). Participants did not receive feedback during the test blocks. The training–test sequence was repeated twice for each participant in the experiment, resulting in 384 training trials and 120 test trials for a total of 504 trials per participant.

3.4. Analysis
All analyses were performed using mixed effect logistic regression (Breslow & Clayton, 1993; Jaeger, 2008) as implemented in the lme4 package (Bates, Maechler, Bolker, & Walker, 2015), with maximal random effects structure supported by the design. That is, the models included random intercepts for participants, and random slopes for VOT and F0 within participants. For all tests, the dependent variable was binary, indicating the participants’ choice of /b/ or /p/. VOT had three values at test: 15, 25 and 35 ms. We are particularly interested in the 15 and 35 ms values because these had a constant frequency across conditions. However, we suspected that a short VOT may be more easily downweighted relative to F0 because F0 becomes available earlier in the speech signal when VOT is short. When VOT is short, the value of VOT becomes available simultaneously with the value of F0 but when it is long, VOT indicates voicelessness a few milliseconds before F0 has a chance to influence perception (see also Toscano & McMurray, 2012). For this reason, we contrast-coded VOT and set the ambiguous VOT value of 25 ms as the reference level for the VOT predictor. In this way, each model included two interactions between condition and VOT. All learning mechanisms predict that F0 should be upweighted in this condition.

4. Results
We examined whether the effects of VOT and F0 on response choice differed across the six conditions in the way predicted by the alternative learning mechanisms. Results are presented in two ways. To facilitate comparison between the prediction of supervised and reinforcement learning to the results obtained from subjects, the results are presented in Fig. 4 below using the same format used for model predictions in Figs. 1 and 2. To provide more detailed information on participants’ choices, the results illustrating participants’ probability of /p/ responses as a function of F0 and VOT in each condition are also presented in Fig. 5.
Supervised and reinforcement learning differ in their predictions for the effect of VOT in condition B. Supervised learning predicts that the magnitude of the effect of VOT in condition B will be equal to its magnitude in condition C, and smaller than in condition A. Reinforcement learning predicts that condition B will pattern with condition A and not C in this respect. That is, downweighting of VOT will be observed only in condition C.

To test these predictions, a logistic regression model with a maximal random effects structure tested the probability of selecting /b/ vs. /p/ as a function of the interaction between VOT and condition, with condition A and VOT of 25 ms serving as the reference levels. Random slopes for VOT within subjects were included. The results indicated a significant interaction between condition and VOT when comparing conditions A and C ($\tilde{\beta} = -0.87, SE = 0.26, z = -3.36, p = .0008$ for VOT 15 vs. 25; $\tilde{\beta} = -0.89, SE = 0.23, z = -3.91, p < .0001$ for VOT 35 vs. 25). However, we found no interaction between condition and VOT when comparing conditions A and B ($\tilde{\beta} = 0.87, SE = 0.26, z = 3.36, p = .0008$ for VOT 15 vs. 25; $\tilde{\beta} = -0.89, SE = 0.23, z = -3.91, p < .0001$ for VOT 35 vs. 25). The BIC approximation to the Bayes Factor indicated that, with the probability of the null hypothesis given the data at 99.99%, we have sufficient power to accept the null hypothesis that, with respect to the effect of VOT, the two conditions A and B do not differ at the population level. The results also indicated a significant interaction between condition and VOT when comparing conditions B and C ($\tilde{\beta} = 0.72, SE = 0.24, z = 3.05, p = .002$ for VOT 15 vs. 25; $\tilde{\beta} = -0.79, SE = 0.21, z = -3.70, p = .0002$ for VOT 35 vs. 25). These results are consistent with the predictions of reinforcement learning but not with those of supervised learning; condition B patterns with condition A rather than condition C.

Both reinforcement learning and supervised learning predict an upweighting of F0 in condition C, and not in condition B. To test this prediction, a logistic regression model with a maximal random effects structure tested the probability of selecting /b/ vs. /p/ words as a function of the interaction between F0 and condition with condition A and F0 of 180 serving as the reference levels. Random slopes for F0 within subjects were included. The results indicated a significant interaction between condition and F0 when comparing conditions A and C ($\tilde{\beta} = 1.53, SE = 0.28, z = 5.37, p < .0001$) as well as conditions B and C ($\tilde{\beta} = 1.43, SE = 0.28, z = 5.07, p < .0001$). There was no significant interaction between condition and F0 when comparing conditions A and B ($\tilde{\beta} = 0.09, SE = 0.09, z = 1.0, p = .32$). The BIC approximation to the Bayes Factor indicated that, with the probability of the null given the data at 98.6%, we have sufficient power to accept the null hypothesis that the two conditions A and B do not differ with respect to the effect of F0. These results are fully consistent with the predictions of both error-driven models.

Supervised learning predicted some downweighting of VOT in the bimodal conditions compared to unimodal conditions, whereas
distributional learning predicted the opposite. To test these predictions, we ran a logistic regression examining the two-way interaction between VOT and modality, which was not significant ($\hat{\beta} = -0.06, SE = 0.20, z = -0.30, p = .76$ for VOT 15 vs. 25; $\hat{\beta} = 0.21, SE = 0.18, z = 1.13, p = .26$ for VOT 35 vs. 25). The BIC approximation to the Bayes Factor indicated that the data provide strong evidence for the null hypothesis of zero effect ($p(H_0|D) = 99.98\%$).

As discussed in the introduction, modality of VOT may interact with modality of F0, so that the listener downweights unimodal VOT only as long as F0 is bimodal, i.e., in condition C. In other conditions, F0 is constant; therefore, the speaker could be inferred to have control over neither F0 nor VOT. To investigate this possibility, another logistic regression model compared specifically the $C_{\text{S}}$ and $C_{\text{M}}$ conditions. There was again no influence of modality on the magnitude of the effect of VOT ($\hat{\beta} = 0.03, SE = 0.37, z = 0.09, p = .93$ for VOT 15 vs. 25; $\hat{\beta} = 0.17, SE = 0.34, z = 0.49, p = .63$ for VOT 35 vs. 25), with strong evidence for the null ($p(H_0|D) = 99.97\%$). There were also no significant interactions between F0 modality (condition C vs. rest) and VOT modality ($\hat{\beta} = 0.10, SE = 0.40, z = 0.24, p = .81$ for VOT 15 vs. 25; $\hat{\beta} = -0.04, SE = 0.36, z = -0.11, p = .91$ for VOT 35 vs. 25; $p(H_0|D) = 100.00\%$) or between informativeness of VOT (condition A vs. rest) and VOT modality ($\hat{\beta} = 0.25, SE = 0.43, z = 0.57, p = .57$ for VOT 15 vs. 25; $\hat{\beta} = 0.27, SE = 0.38, z = 0.71, p = .48$ for VOT 35 vs. 25; $p(H_0|D) = 99.98\%$). The absence of any modality effects on the weight of VOT is inconsistent with both distributional learning and supervised learning. In contrast, it is consistent with the reinforcement learning model.

Finally, supervised learning predicted an effect of modality on the weight of F0 in the C condition: F0 was expected to be upweighted more in the $C_{\text{M}}$ condition compared to the $C_{\text{S}}$ condition. The opposite prediction was made by distributional learning. Reinforcement learning again predicted no effect of modality. To test these predictions, we examined the difference between the effect of F0 in the $C_{\text{M}}$ condition and in the $C_{\text{S}}$ condition, by fitting a two-way F0 × Modality interaction within the C condition. This comparison was not significant ($\hat{\beta} = -0.38, SE = 0.61, z = -0.62, p = .53$), with strong evidence for the null hypothesis ($p(H_0|D) = 98.58\%$). We also ran a logistic regression model with a three-way interaction between modality of VOT, modality/informativeness of F0 (condition A and B vs. C) and F0 value. The Rescorla–Wagner model predicted this interaction to be significant because modality is expected to influence the weight of F0 only in condition C, where F0 is upweighted. We found no significant interaction ($\hat{\beta} = -0.09, SE = 0.38, z = -0.24, p = .81$; with strong evidence for the null ($p(H_0|D) = 99.30\%$)). These results are inconsistent with both supervised learning and distributional learning, but are consistent with reinforcement learning.

The reported results above are aggregated across the two test blocks, which were separated by a repetition of the training block (in a different random order of trials). The results were qualitatively unaffected by block, but the significant interactions became stronger with additional training. Thus, the interaction between condition (C vs. rest) and F0 increased from test block 1 to test block 2 ($\hat{\beta} = 0.94, SE = 0.15, z = 6.19, p < .0001$), although it was also significant during block 1 ($\hat{\beta} = 0.90, SE = 0.19, z = 4.71, p < .0001$). Similarly, the interaction between condition and VOT was significant in block 1 ($\hat{\beta} = 0.74, SE = 0.23, z = 3.18, p = .0015$ for VOT 15 vs. 25; $\hat{\beta} = -0.60, SE = 0.21, z = -2.89, p = .004$ for VOT 35 vs. 25) but the downweighting of VOT increased in magnitude in test block 2, at least for the 25 ms vs. 35 ms contrast ($\hat{\beta} = -0.50, SE = 0.20, z = -2.45, p = .014$). It is therefore possible that continuing training would result in continued upweighting of F0 and downweighting of VOT.

Overall, the results show downweighting of VOT and upweighting of F0 in condition C, but not in condition B, and no effect of VOT distribution modality. This pattern of results is consistent only with the predictions of the reinforcement learning model.

5. Discussion

We tested alternative views of the influence of prediction error and distributional information on the weighting of cues to a familiar phonological contrast, voicing. For a subset of participants, F0 was a perfect predictor of voicing categories while for others it was constant during training, thus providing no information regarding whether variation in F0 is predictive of voicing. Learners significantly decreased reliance on VOT and increased reliance on F0, the secondary cue in their language, when and only when they could rely on F0 to successfully predict voicing during training.

5.1. Supervised learning

The supervised Rescorla–Wagner model (Rescorla & Wagner, 1972; see also Arnold et al., 2017; Baayen, Milin, Durdević, Hendrix, & Marelli, 2011; Olejarzuk et al., 2018; Ramscur, Yarlett, Dye, Deny, & Thorpe, 2010, 2013a, 2013b) successfully predicted that F0 would be upweighted only in condition C, where F0 is predictive of feedback. However, in other ways, the results are highly discrepant with the predictions of the Rescorla–Wagner model. First, the Rescorla–Wagner model expected equivalent downweighting of VOT regardless of whether F0 was informative about voicing. As a result, the Rescorla–Wagner model predicted that there should be a fairly large difference between condition A, where VOT is informative about voicing, and condition B, where it is uninformative. In contrast, Bayesian analysis of our data provided strong evidence for the hypothesis that there was no effect of VOT informativeness. Our participants do not appear to be using prediction error in the way we would expect from the Rescorla–Wagner model.

Second, the Rescorla–Wagner model predicted that more learning about VOT and F0 should occur when the VOT distribution is bimodal than when it is unimodal. Like the prediction of a difference between conditions A and B, this is a direct consequence of the use of prediction error in the Rescorla–Wagner model. Reliance on VOT generates more error when VOT is distributed bimodally. When the distribution is unimodal, many training trials feature ambiguous VOT (25 ms). On these trials, equal $p(l/-/b/)$ proportions are expected on the basis of the participants’ prior experience. Therefore, feedback on these trials is consistent with the participants’ prior beliefs and does not result in prediction error. In contrast, 50/50 feedback after responses to stimuli with unambiguous VOT values (5 ms and 45 ms) results in much prediction error and should therefore greatly shift the participants’ beliefs. Error-driven supervised models like Rescorla–Wagner necessarily lead us to expect more learning in the bimodal condition, contrary to the results.

5.2. Reinforcement learning

In contrast, the results are fully consistent with reinforcement learning, which predicts that the learner reallocates attention across perceptual dimensions when she estimates that doing so would reduce prediction error on future trials. Switching attention from VOT to F0 in response to prediction error caused by reliance on VOT would not result in a decrease in prediction error in condition B, just as it would not result in such a decrease in condition A. The lack of learning in condition B is therefore expected based on reinforcement learning. More generally, reinforcement learning leads us to expect that learners will persist with previously reinforced actions until the environment convinces them that another behavior would be more beneficial, at which point a rapid shift in policy would be expected (e.g. Gallistel, Fairhurst, & Balsam, 2004; Silberberg, Hamilton, Zriak, & Casey, 1978).

While reinforcement learning necessarily predicts that condition B will pattern with condition A, and not with condition C, other patterns in the data are captured by the reinforcement learning model only because of certain modeling choices. First, we treated error as...
categorical, i.e., as either predicting the right output or not. The categorical nature of error in the model assumes that a correct prediction is equally reinforcing, regardless of how confidently it was made. Similarly, the model assumes wrong prediction is equally punishing, regardless of how confidently it was made. Alternatively, it could be assumed—along the lines of the Rescorla–Wagner model—that less confident predictions are reinforced more strongly when they are confirmed, and disconfirmations of confident predictions are more punishing (i.e., error is gradient). No upweighting of F0 is expected by the reinforcement learning model with gradient error in condition C; only downweighting of VOT is expected. We did not find this a plausible prediction, and our data disconfirm it: there is significant up-weighting of F0 in condition C.

As noted in the introduction, a second crucial choice in setting up the reinforcement learning model was that only the policies of maximum value were allowed to influence behavioral choices. An alternative is to allow all policies to drive choices, with the influence of each policy proportional to its value. However, this results in the reintroduction of the difference between the Cuni and Cbi conditions predicted by the Rescorla–Wagner model. The original, English-like weights make fewer errors in the Cuni condition because the random assignment of VOT values to phonological categories is consistent with English when these values are in the middle of the VOT continuum. As a result, the original policy, and other policies like it, are downweighted less strongly in the Cuni condition. They will therefore exert more influence on behavior unless excluded from the set of influential policies. Maximizing value prevents them from exerting any influence on behavior, eliminating the difference between the Cuni and Cbi conditions.

5.3. Distributional learning

An alternative perspective on phonological cue weighting is provided by distributional learning (Feldman et al., 2009; Kleinschmidt & Jaeger, 2015; Maye & Gerken, 2000; Maye et al., 2002). As discussed earlier, a unimodal distribution along VOT could lead a generative learner to infer that the speaker has a single VOT target in production (e.g., Feldman et al., 2009; Kleinschmidt & Jaeger, 2015) and therefore cannot be using VOT to convey semantic distinctions. We hypothesized that this inference may then lead the learner to shift attention away from VOT. As we discuss below, there is a wide space of distributional learning models whose predictions may vary, except for one constant: barring floor effects, the Cuni condition should produce more down-weighting of VOT than the Cbi condition. The absence of this effect in the present study is therefore problematic for distributional learning.

From a distributional learning perspective, the lack of learning or VOT modality effects in condition B might be explained by the distributional learner's sampling assumptions, i.e., beliefs about how the sample of productions they encountered was generated from the underlying population of productions. We intended the learners to take the constant F0 as indicating lack of information regarding how the speaker uses variation in F0. In that case, the learners may be expected to downweight VOT, which they know to be unimodally distributed in the speaker's idiolect, relative to F0, about which they have no information. However, this interpretation of the training may require the sampling assumption that the experimenter has controlled F0. Instead, learners could assume weak sampling, considering the training examples to be a random sample from the speaker's productions (e.g., Navarro, Dry, & Lee, 2012; Xu & Tenenbaum, 2007). The constancy of F0 should then be interpreted as indicating that the speaker does not vary F0 in production, and should therefore prevent attending to F0. Under the assumption of weak sampling, reduction in VOT weight may therefore be expected to occur only in condition Cuni where F0 is distributed bimodally.

Distributional learning may also be expected to interact with error-driven learning (e.g., Escudero, 2005; Hayes-Harb, 2007). Speakers in the Bbi and Cbi conditions appear to have two production targets (phones) for /b/ and /p/ but select them randomly regardless of referential intention. That is, speakers in Bbi and Cbi appear not to know which phone occurs in which word. Previous research has suggested that listeners can rapidly adjust word-phone mappings from perceptual experience based on lexical feedback (German, Carlson, & Pierrehumbert, 2013). If learners build a generative model of the speaker, that model should therefore include not only the acoustic distributions within phonological categories but also the mappings between phonological categories and the words that contain them (Escudero, 2005; German et al., 2013). Therefore downweighting of VOT in the Cbi condition per se does not necessarily challenge distributional learning: it might simply result from learning word-phone mappings instead of the mappings between phones and locations in the VOT/F0 space that distributional learning is concerned with.

However, barring a floor effect, distributional learning necessarily predicts greater downweighting of VOT and upweighting of F0 in the Cuni condition compared to Cbi. As discussed above, this prediction contrasts with the predictions of error-driven learning. In our data, the weight of VOT is far from floor in condition C, with VOT continuing to exert a significant influence on the participants' choice behavior. Nonetheless, we see no effect of modality, which is problematic for distributional learning.

An additional problem for distributional learning models is the lack of a modality effect in condition A. The voicing categories in condition A are largely English-like: the speaker encountered in this condition has bimodal distributions along both VOT and F0, and English-like mappings between phones and locations in the F0–VOT space. In contrast, condition Auni is highly unexpected from a distributional perspective. Assuming that production tokens are normally distributed around the target (Kleinschmidt & Jaeger, 2015), the Auni speaker appears to have only one production target along VOT but maps the VOT values to the left of the mode onto /b/ and those to the right of the mode onto /p/. That is, lexical feedback conflicts with distributional information in this condition. Yet, the behavior of participants exposed to the Auni and Abi conditions is remarkably similar.

One possible explanation is that the presence of feedback in our study may have made distributional learning less efficacious than it would otherwise be. A rather substantial amount of experienced data is necessary to reliably infer the modality of a distribution (Wanrooj et al., 2015). Quite a bit less data is necessary to change beliefs about the most likely dimensional weights. This might make distributional learning weaker when the learner can take advantage of feedback and prediction error. However, prior studies have had difficulty demonstrating clear evidence of distributional learning even without feedback (Hayes-Harb, 2007; Wanrooj, Boersma, & van Zuijen, 2014). It may therefore be the case that distributional learning is unavailable for previously categorized dimensions (Escudero, 2005).

6. Conclusion

How do we adjust our beliefs when the world changes? The present paper contrasted the predictions of two prominent views of learning from prediction error—supervised learning (Rescorla & Wagner, 1972) and reinforcement learning (Sutton & Barto, 1998)—in the domain of revising the weights of perceptual cues to phonological (or lexical) categories, a crucial speech perception task. Our results contradict several predictions inherent to the Rescorla–Wagner model of supervised error-driven learning. Most importantly, supervised learning led us to expect that prediction error caused by reliance on a cue should lead learners to downweight this cue even when there is no other cue to rely on. We observed that downweighting of a cue instead occurs only if relying on another cue reduces prediction error. While inconsistent with supervised error-driven models in the *Rescorla–Wagner (1972) tradition, this result is expected from a reinforcement perspective on error-driven learning (Sutton & Barto, 1981, 1998).

By providing support for reinforcement learning, these results
bolster the contribution of domain-general error-driven learning mecha-
nisms to language acquisition (see also Ellis, 2006; Francis & Nusbaum, 2002; Holt, Lotto, & Kluender, 1998; Kapatsinski, 2018; Ramscar et al., 2013a) while elucidating the nature of these mechan-
isms. Even though supervised models based on the Rescorla–Wagner rule have enjoyed widespread success in accounting for surprising phenomena in language acquisition (e.g. Arnold et al., 2017; Baayen et al., 2011, 2016; Olejarczuk et al., 2018; Ramscar et al., 2013a; Ramscar, Dye, & McCauley, 2013b; Ramscar et al., 2010, 2014), the present results suggest that language learners may not blindly down-
weight cues when faced with prediction error but only do so when this change in policy is expected to reduce prediction error in the future.

7. Research data and models for this article

Data, statistical analyses, and model code are available at: https://
osf.io/v576c/.

Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. We are grateful to two anonymous reviewers and T. Florian Jaeger whose comments on an earlier draft of this paper have considerably improved the paper. This work has been presented at Interdisciplinary Advances in Statistical Learning 2017, 173rd Acoustical Society of America Meeting, Boston University Conference on Language Development (BUCLD 42), and 58th Annual Meeting of the Psychonomic Society. We thank the audience at these conferences for helpful feedback.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2019.03.011.

References

miliar, generalize to the similar, and adapt to the novel. Psychological Review, 122(2), 148–203.
ers to identify English /b/ and /d/. In S. Y. Tolkia, E. Vatikiotis-Paterson, & Y. Sagisaka (Eds.). Speech perception, production and linguistic structure (pp. 175–196). Amsterdam: IOS Press.
Somerville, MA: Cascadilla Press.


