COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 48 (2024) e13450

ISSN: 1551-6709 online
DOI: 10.1111/cogs.13450

One Cue's Loss Is Another Cue's Gain—Learning Morphophonology Through Unlearning

Erdin Mujezinović,[a] Vsevolod Kapatsinski,[b] Ruben van de Vijver[a]

[a]*Institute for Linguistics, Heinrich-Heine University Düsseldorf*
[b]*Department of Linguistics, University of Oregon*

Received 25 September 2023; received in revised form 3 April 2024; accepted 13 April 2024

Abstract

A word often expresses many different morphological functions. Which part of a word contributes to which part of the overall meaning is not always clear, which raises the question as to how such functions are learned. While linguistic studies tacitly assume the co-occurrence of cues and outcomes to suffice in learning these functions (Baer-Henney, Kügler, & van de Vijver, 2015; Baer-Henney & van de Vijver, 2012), error-driven learning suggests that contingency rather than contiguity is crucial (Nixon, 2020; Ramscar, Yarlett, Dye, Denny, & Thorpe, 2010). In error-driven learning, cues gain association strength if they predict a certain outcome, and they lose strength if the outcome is absent. This reduction of association strength is called unlearning. So far, it is unclear if such unlearning has consequences for cue–outcome associations beyond the ones that get reduced. To test for such consequences of unlearning, we taught participants morphophonological patterns in an artificial language learning experiment. In one block, the cues to two morphological outcomes—plural and diminutive—co-occurred within the same word forms. In another block, a single cue to only one of these two outcomes was presented in a different set of word forms. We wanted to find out, if participants unlearn this cue's association with the outcome that is not predicted by the cue alone, and if this allows the absent cue to be associated with the absent outcome. Our results show that if unlearning was possible, participants learned that the absent cue predicts the absent outcome better than if no unlearning was possible. This effect was stronger if the unlearned cue was more salient. This shows that unlearning takes place

Correspondence should be sent to Erdin Mujezinović, Institute for Linguistics, Heinrich-Heine University Düsseldorf 1, 40225 Düsseldorf, Germany. E-mail: Erdin.Mujezinovic@hhu.de

even if no alternative cues to an absent outcome are provided, which highlights that learners take both positive and negative evidence into account—as predicted by domain general error-driven learning.

Keywords: Error-driven learning; Discriminative learning; Unlearning; Negative evidence; Overshadowing; Learnability problem; Morphophonology; Phonology

1. Introduction

What evidence do humans rely on to learn a language? In theoretical linguistic work, it is often assumed that only positive evidence is available for learning but not negative evidence. What counts as negative evidence, however, is also not discussed in detail, and, to the extent it is discussed at all, there is no consensus about what would constitute negative evidence. Chouinard and Clark (2003), for example, assume that recasts, and Albright and Hayes (2011) that explicit corrections are negative evidence, while Yang (2016) more broadly claims that not only overt corrections count as negative evidence but also constructions that a learner never encounters. Interestingly, for Yang such evidence should be interpreted by learners as evidence that these constructions are ungrammatical. It is unclear if learners do interpret such absence of evidence as evidence of absence (Bowerman, 1988; Marcus, 1993). Linguistic theories thus in general refrain from discussing negative evidence, focusing on positive evidence only.

This view of evidence is in line with a theory of language learning, in which learning consists of establishing weights between two events when these co-occur in close temporal proximity (Bybee, 2010; Maye, Werker, & Gerken, 2002). The more often a cue and an outcome co-occur, the stronger their association becomes. As a result, the knowledge of learners reflects the distribution of cues and outcomes in their input (Ambridge, Kidd, Rowland, & Theakston, 2015; Bybee, 1995; Mirković, Seidenberg, & Joanisse, 2011). Although much evidence supports the view that learning is affected by this kind of positive evidence (Adriaans & Kager, 2010; Ambridge et al., 2015; Baer-Henney, Kügler, & van de Vijver, 2015; Baer-Henney & van de Vijver, 2012; Hayes, Zuraw, Siptár, & Londe, 2009; Hayes & Londe, 2006; Kapatsinski, 2010; Olejarczuk & Kapatsinski, 2018; Saiegh-Haddad, Hadieh, & Ravid, 2012; Song & White, 2022; Szagun, 2011), the role of negative evidence is left unexplored.

The present study addresses the role of negative evidence by investigating the learning mechanisms used during morphophonological learning. Our aim is to provide insights about whether learners make use of negative evidence and in what way both positive and negative evidence contribute to form-meaning associations. We will first describe the central tenets behind learning theory and error-driven learning before moving on to research in language learning and the objectives of our study.

1.1. Learning theory

The learning mechanisms responsible for learning cue–outcome associations were first addressed in research that investigated how animals learn (Kamin, 1967; Rescorla, 1988; Rescorla and Wagner, 1972). This work showed that associations are not solely based on

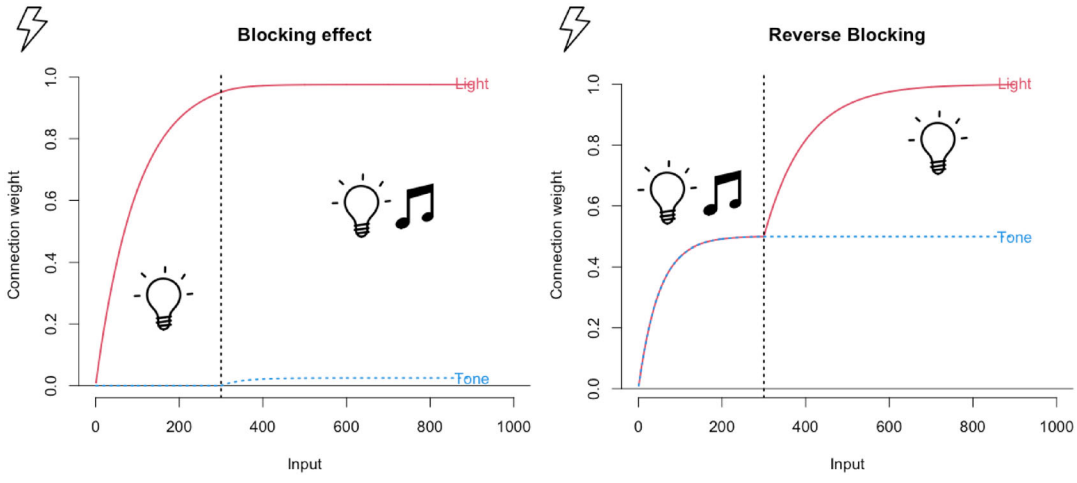

Fig. 1. Error-driven learning as highlighted by the blocking effect. The connection weights show the association to the shock outcome for the light cue (red) and the tone cue (blue). The dashed line marks the beginning of the input change. If light-only input is encountered first (left plot), the tone-shock association is not learned. If light and tone together are encountered first (right plot), tone is also associated with the shock, although to a lesser degree than light. The simulation is based on the Rescorla–Wagner learning equations (see Section 3.1), using `edl` (van Rij & Hoppe, 2021).

contiguity but on predictions. If a cue helps to predict an outcome, their association is strengthened, and if the cue does not help it is weakened. Crucially, the associations are updated in an error-driven way; an association is more strongly affected if there is a greater amount of error. The importance of prediction and prediction error is best shown by the *blocking effect*, in which an association between a co-occurring cue and an outcome association is not learned (Kamin, 1967). Kamin (1967) first taught rats that a flash of light was followed by a mild electric shock and, subsequently, that a flash of light together with a tone was followed by a mild electric shock. In a third step, Kamin tested whether the rats had associated both light and tone with the mild shock. He found that the rats had only learned the association between light and shock but not between tone and shock. Kamin's explanation is that learning is predictive; based on contingency rather than contiguity of events. Although the tone also co-occurred with the shock, there was no additional role for the tone in predicting it, as the presence of a light was sufficient. Because the tone had no predictive value, the association was not learned. When the order of learning was reversed, so that light together with tone was experienced first, the blocking effect did not occur, and the rats associated both light and tone with the shock (Kamin, 1967). How the association between the cues and the shock develops is illustrated in Fig. 1. This kind of learning is known as error-driven learning (Hoppe, Hendriks, Ramscar, & van Rij, 2022); (Kapatsinski, 2018b, 2023a; Nixon, 2020; Olejarczuk, Kapatsinski, & Baayen, 2018; Ramscar, Yarlett, Dye, Denny, & Thorpe, 2010; Rescorla & Wagner, 1972).

How do association weights translate to observable behavior? Miller and Matzel (1988) proposed the *comparator hypothesis*, which assumes that associations developed between cues and an outcome are compared to each other (Denniston, Savastano, & Miller, 2000; Miller & Matzel, 1988; Stout & Miller, 2007). Consider the scenario of reverse blocking: Even though the tone is associated with the shock, this association is weaker than the association between light and shock. Because of this, a learner would be more likely to expect a shock upon light exposure than a shock upon tone exposure. This means that not the raw association strength, but the relative association strength matters, so that the association weights of other cues may also influence learning.

1.2. Error-driven learning

The blocking effect shows that not every co-occurrence of a cue–outcome association influences learning to the same degree. Within error-driven learning, cues predict outcomes, so that an association from cue to outcome gets strengthened each time the cue predicts the outcome. This means that if cues that more reliably predict an outcome are available, other cues will not be learned, even if they are present. Thus, what sets error-driven learning apart from associative learning is that cue–outcome co-occurrences are not tracked in isolation, but that all present cues compete with each other for their share of activation during a learning event. This is called *cue competition* (Hoppe et al., 2022; Nixon, 2020; Ramscar et al., 2010; Rescorla & Wagner, 1972). Activation from a cue to an outcome increases in proportion to the prediction error present during a learning event (Nixon & Tomaschek, 2023). If a cue often predicts an outcome, there is less uncertainty about whether the outcome follows, and, as a result, there is also less prediction error remaining. This cue takes up all activation for itself, leaving no share to other cues. This explains why additional cues are sometimes not learned (Kamin, 1967).

What happens if the predictive cue is present, but a previously encountered outcome is absent? When a previously encountered outcome is absent, the learner encounters a prediction error, which results in a decrease in association strength. Learners, therefore, track both the occurrence and non-occurrence of outcomes (Hoppe et al., 2022). This ensures that non-predictive cues become irrelevant, while predictive ones get strengthened. It also ensures that cues that are not predictive anymore become dissociated from an outcome, for example, when circumstances change, for example, when a person changes their last name so that this person is not predictive of the old name anymore. Learners have to not only associate that person with the new name, but the old name also has to get dissociated. This effect is called *unlearning* (Nixon, 2020). In error-driven learning, positive and negative evidence and prediction and prediction error interact and shape the learning process. Computationally, these learning mechanisms are captured in a two-layer neural network that updates association weights based on the Rescorla–Wagner equations (Rescorla & Wagner, 1972). In essence, association strengths of all present cues to an outcome are increased, if the outcome is present (positive evidence), and they are decreased, if the outcome is absent (negative evidence). We will provide further details on the Rescorla–Wagner equations and their computational implementation in Section 3.1.

1.3. Error-driven learning in language

Error-driven learning is a domain of general learning theory that has been applied to many topics in cognition (Hoppe et al., 2022) and has recently been applied to language (Baayen, Chuang, & Blevins, 2018; Baayen, Chuang, Shafaei-Bajestan, & Blevins, 2019; Baayen, Hendrix, & Ramscar, 2013; Baayen, Milin, Đurđević, Hendrix, & Marelli, 2011; Baayen, Shaoul, Willits, & Ramscar, 2016a; Chuang et al., 2021; Denistia & Baayen, 2023; Nieder, Tomaschek, Cohrs, & van de Vijver, 2021; Nieder, Chuang, van de Vijver, & Baayen, 2023; van de Vijver & Uwambayinema, 2022; van de Vijver, Uwambayinema, & Chuang, 2024), and language learning (Divjak, Milin, Ez-zizi, Józefowski, & Adam, 2021; Ellis, 2006; Harmon, Idemaru, & Kapatsinski, 2019; Nixon, 2020; Olejarczuk et al., 2018; Ramscar, Dye, & McCauley, 2013; Ramscar & Yarlett, 2007; Ramscar et al., 2010; Romain, Ez-zizi, Milin, & Divjak, 2022).

In error-driven learning, it is possible that a cue is present in the input, but that it does not impact learning, as it does not affect prediction error. A linguistic example is provided by Nixon (2020), who investigated whether the blocking effect Kamin (1967) observed in rats also occurs in language learning. She did so by investigating how the associations between alien pictures and phonetic features of pseudowords, used to refer to these aliens, were learned. In one experiment, she first taught participants that a word form with a lexical high tone predicted an alien. In the second step, the alien was predicted by a lexical high tone and a nasalized vowel. In a third step, she tested whether participants had associated nasality with the alien and found that they had not. This result shows that blocking also occurs in language learning; evidence is ignored under certain circumstances.

The blocking effect appears to be general and has also been found in artificial language learning experiments on noun class learning. Culbertson, Gagliardi, and Smith (2017) investigated whether learners rely more on semantic cues or phonological cues to noun classes. They taught one group of participants that semantic cues predicted particular word classes in a first step, followed by a second step, where they taught them that phonological cues predict those word classes as well. This order was reversed for a second group of participants. In the test blocks, they found that the first group had only associated semantic cues with noun classes, and the second group had only associated phonological cues with noun classes. Even though Culbertson et al. explain this as an effect of salience and availability of cues, resulting from the order of presentation, their finding can more generally be explained as an effect of blocking.

Blocking effects have also been attested in second language acquisition studies. When learning verbal tense morphology, Ellis and Sagarra (2010a) have found that previous knowledge about a morphological construction that conveys this meaning can block the learning of other constructions that convey the same meaning. When participants were pre-trained with Latin adverbs that convey temporal relations, for example, *heri* "yesterday," they did not learn verbal morphology, for example, *-avi* in *cogitavi* "I thought," and when they were trained with verbal morphology they did not learn the temporal adverbs. Similar findings have been found in different languages and with different methods, for example, eye-tracking (Ellis & Sagarra, 2010b; Ellis et al., 2014; Sagarra & Ellis, 2013). Previously learned contingencies, therefore, can block additional learning of contingencies that express the same function.

Over time, evidence can also lose its impact. When a cue is present while an expected outcome is absent, the link between them is weakened. When this happens repeatedly, the cue–outcome association is unlearned. This unlearning effect was shown by Ramscar et al. (2010) and Nixon (2020). Ramscar et al. (2010) showed that novel labels for semantic categories are better learned, if more salient—but uninformative—body shape cues are unlearned. Likewise, in addition to finding a blocking effect, Nixon (2020) has also shown that using lexical tones to distinguish between geometrical forms is better learned, if more salient—but uninformative—syllable shapes are unlearned. Learning both novel labels and geometrical forms, therefore, involved an increase in association for informative and a decrease in association for uninformative cues. Learners may not only ignore evidence which is present, but they may also learn from evidence which is absent.

That both positive and negative evidence contribute to learning about contingencies has also been demonstrated for the learning of English plurals. Ramscar et al. (2013) use error-driven learning to address the problem of learning irregular English plural words. Many English plurals are regular and expressed by the suffixes [s], [z], or [əz], and some are irregular and expressed by vowel changes as in *mouse˜mice*, and a handful of other endings, such as [ən] in *oxen*. Children typically go through a phase, in which they make irregular plurals regular, but hardly ever generalize irregular forms to novel word forms (Marcus et al., 1992; Pinker, 1984; Prasada & Pinker, 1993; Yang, 2016). Ramscar et al. (2013) ask how they can retreat from this erroneous pattern if only positive evidence is available. Their answer is couched in the way error-driven learning allows for negative evidence to influence learning (Rescorla & Wagner, 1972): A child will hear adults use regular plurals, such as *cats, dogs, horses*, in the context of actual CATS, DOGS, and HORSES more often than irregular plurals, such as *mice*, in the context of actual MICE. At first, this will lead the child to expect that a plural meaning predicts a word form ending in a sibilant—and if the word already ends with a sibilant, in a word form ending with an [əz]. The child, therefore, produces **mouses* instead of *mice*. However, the child will observe actual MICE and hear the word form *mice*—and not the final [əz]—and this prediction error leads her to downweight the association between PLURAL and [əz]. Over time, this results in a pattern that looks like a retreat from an erroneous pattern, which in reality is caused by a change in connection weights as a result of prediction errors. The evidence necessary for learning is present in the input, and the evidence is partitioned into positive and negative evidence to the extent that a cue correctly or incorrectly predicts a word form. Still, to unlearn the non-predictive association between MICE and *mouse* + final [əz], learners also need to encounter other informative cues with the outcome, that is, MICE and *mice*, to retreat from overregularizations.

That additional informative cues that are necessary for such unlearning to take place have recently been shown by Harmon et al. (2019) for phonetic category learning. English speakers had to unlearn their previous association between voicing and VOT—the primary cue for voicing contrasts in stop sounds in English—and instead learn to associate voicing with F0—a secondary cue for that contrast in English. They found that English speakers only unlearned the VOT association if the secondary F0 cue was informative instead. If F0 was also uninformative, English speakers continued to rely on their prior associations, even when they were not helpful anymore. Harmon et al. conclude that for unlearning to

take place, encountering an error is not enough: Unlearning has to actually reduce future errors. The circumstances under which unlearning takes place, therefore, still need further investigation.

1.4. When and how does unlearning come about?

According to error-driven learning theory (Rescorla & Wagner, 1972), when a tone cue predicts a food outcome, this association will be unlearned as soon as the tone no longer predicts the food. Importantly, unlearning only takes place in cases where a cue is present, while an outcome is missing: If the tone cue is absent, nothing is learned. We do not know yet if unlearning always occurs under this condition, and in what way it affects other learned associations.

First, there is evidence that unlearning only occurs if other, more reliable cues are available (Harmon et al., 2019). Harmon et al. point out that the Rescorla–Wagner equations update associations only in accordance with the presence of prediction error. Whether this results in a reduction of future errors is not taken into account. An association is either strengthened—learned—if a cue predicts an outcome or reduced—unlearned—if the outcome fails to occur. According to Harmon et al., the Rescorla–Wagner equations act as if they are blind to the consequences of their update. Because of this, they are not compatible with Harmon et al.'s finding that unlearning will only take place if positive evidence for the informativity of other cues to the same outcome is also available.

Second, it is not clear yet what happens when learners do not know whether another cue would really predict an outcome, which is not predicted by a present cue. This happens for, instance, if at first two cues predict not only one but two outcomes simultaneously, and, in the next step, one of the cues predicts only one of these outcomes. Does the presence of this cue, while the other outcome is absent, lead to their association being unlearned? And does this take place even when learners do not encounter conclusive positive evidence for the association between absent cue and absent outcome? If, for example, both a light and a tone cue have been encountered with a food outcome first, and then the tone cue is encountered with a different outcome (e.g., a toy), the tone–food—but not the light–food association—will be affected. Even though associations are only updated when cues are present, such unlearning is also likely to impact the light–food association in an indirect way: We hypothesize that the tone will interfere less with this association given its decreased association. This would then result in a stronger association between light cue and food outcome. If unlearning follows under this circumstance, learners get two for the price of one: they can learn the association between a present cue and its present outcome but also about a previously encountered—but currently absent—cue, and its previously encountered—but currently absent—outcome. While such an effect is reminiscent of the retrospective revaluation sometimes observed in causal and animal learning (Blaser, Couvillon, & Bitterman, 2004; Chapman, 1991; Miller & Witnauer, 2016; Van Hamme & Wasserman, 1994), the latter effects involve an update of association weights based on absent cues: In retrospective revaluation, learning that two cues predict one outcome, and subsequently that only one of these cues predicts this outcome results in the other cue losing its association to this outcome. The former effect though involves updating

associations only when cues are actually present, with the association between absent cue and absent outcome emerging as a consequence of this.

Would unlearning from negative evidence occur and affect other previously learned cue–outcome associations? We argue that as long as there is no conclusive evidence that other cues are uninformative, unlearning will still take place. Unlearning will only fail to occur, if no informative cues to the same outcomes are available, as was the case in Harmon et al. (2019). In such situations, learners will hold on to previously learned, but non-predictive associations. In cases where other cues could potentially be informative, unlearning would be possible, even when no positive evidence is presented explicitly. Thus, if new evidence resolves an ambiguity between cues and outcomes, as outlined above, unlearning would result in not only a reduction of error for a present cue and a present outcome but also for previously encountered absent cues and absent outcomes, as the latter associations can benefit from the reduced weight of the former ones. This would show that both positive and negative evidence have a profound effect on learning.

1.5. The present study

The present study explores the role of unlearning for ambiguous stimuli in the domain of morphophonology. We do this by means of an artificial language learning experiment. Phonological word forms served as cues that predicted morphological outcomes, in line with previous studies on phonetic and phonological learning (Harmon et al., 2019; Nixon, 2020; Nixon & Tomaschek, 2021; Olejarczuk et al., 2018). We are interested in whether positive evidence only affects learning, or if negative evidence is additionally used, as it is still unclear, what kinds of evidence learners use to form linguistic generalizations. By testing morphophonological learning, we ensure that any consequences of unlearning observed in our study follow from learning within the experiment. Our study contributes to the understanding of negative evidence in language learning (Goldberg, 2019; Pearl, 2022; Ramscar et al., 2013; Yang, 2016) by testing whether implicit negative evidence impacts language learning, as well as how such effects may emerge from domain general learning mechanisms. We will first describe our methods in Section 2, followed by the computational simulations in Section 3 and the learning experiment in Section 4.

2. Methods

2.1. The artificial language

We prepared an artificial language which featured word forms for singular, plural, diminutive, and diminutive-plural nouns of a morphological paradigm. We modeled the morphological pattern after the i-declension of masculine nouns of the Serbo-Croatian language (Kordić, 1997), as we needed a pattern which German speakers were unlikely to know. We chose Serbo-Croatian because it was the native language of the first author. Serbo-Croatian marks plural and diminutive nouns with distinct suffixes, both of which are present in diminutive-plural nouns. In addition, singular word forms contain a distinct phonological marker within

Table 1
Structure of the stimuli in the artificial language learning experiments

Category	C_1	V	C_2	C_3	Cue	Example
Singular	{b, d, g, z}	{e, o, u}	{p, t, k, f}	{m, n, l}	[a]	[be.fan]
Plural	{b, d, g, z}	{e, o, u}	{p, t, k, f}	{m, n, l}	[i]	[bef.ni]
Diminutive	{b, d, g, z}	{e, o, u}	{p, t, k, f}	{m, n, l}	[it͡ʃ]	[bef.nit͡ʃ]
Diminutive-plural	{b, d, g, z}	{e, o, u}	{p, t, k, f}	{m, n, l}	[it͡ʃi]	[bef.ni.t͡ʃi]

the chosen paradigm as singular word forms include a word-medial [a], which is absent in other word forms. This ensures a direct mapping from parts of a word form to all four morphological categories. The structure of the artificial language is shown in Table 1.

We constructed the stimuli by carefully controlling for each item's phonological structure. We used a subset of the intersection between the Serbo-Croatian and German phoneme inventories (Kordić, 1997; Subotić, Sredojević, & Bjelaković, 2012; Wiese, 1996). The nominal root of all words consisted of a $C_1VC_2C_3$ pattern, as exemplified by the root /befn/. The singular word forms included an [a] infix between C_2 and C_3—reminiscent of the vowel-zero alternation in Serbo-Croatian (Kordić, 1997), for example, [be.fan]. Plural and diminutive nouns had an [i] suffix (e.g., [bef.ni]) and an [it͡ʃ] suffix (e.g., [bef.nit͡ʃ]). For diminutive-plurals both suffixes co-occurred, the diminutive suffix being closer to the root (e.g., [bef.ni.t͡ʃi]). The word forms were subject to resyllabification, in cases where illicit consonant clusters in German would otherwise occur. We varied the sounds of the $C_1VC_2C_3$ roots in a systematic way: C_1 was always a voiced obstruent {b, d, g, z}, V was always a vowel {e, o, u}, C_2 a voiceless obstruent {p, t, k, f}, and C_3 a sonorant {m, n, l}. We excluded voiced obstruents in the C_2 position to preempt complications arising from final obstruent devoicing in German (Wiese, 1996). C_2 occurs in the coda position in all but the singular word forms, so that final devoicing would change the obstruents to their voiceless counterparts. We hoped to rule out stimuli-specific factors that may affect learning by controlling for the phonological structure.

To arrive at the final item list, we first concatenated every possible sound combination for $C_1VC_2C_3$ roots, which resulted in a total of 144 stems ($4 \times 3 \times 4 \times 3$). We then constructed the actual word forms for each category. We removed items that resembled actual German words (e.g., [dekan] "dean") as well as items that violated German phonotactics. This resulted in 107 different word roots. Finally, we trimmed this set further down to 81 roots by removing minimal pair members that differed in the final nasal sound. We did this because the nasal resonances make the formant transitions from vowel to nasal—which are responsible for place distinction—hard to hear (Zsiga, 2020, p. 222). Participants are thus likely to confuse [m] and [n]. We alternated between removing the labial and the alveolar nasal of a pair. For familiarization, we used the /befn/ root, that is, the word forms [be.fan] (singular), [bef.ni] (plural), [bef.nit͡ʃ] (diminutive), and [bef.ni.t͡ʃi] (diminutive-plural). We pseudorandomized the other 80 roots first and then divided them into three item sets (learning block I, learning block II, and test block only items). Twenty-seven roots contained a close back vowel [u], 27 a close-mid back vowel [o], and 26 a close-mid front vowel [e]. Each word initial obstruent {b,

d, g, z} occurred equally often (4 × 20). Learning blocks I and II contained 32 roots each, while the test block only set contained 16 roots. In total, 224 distinct word forms were used during the learning part of the experiment: 64 singulars, plurals, and diminutives each, as well as 32 diminutive-plurals. The full itemlists are shown in the Supporting Information (see Tables S5–S10).

2.2. The design of the learning experiment

We prepared two conditions in which we tested the consequences of unlearning for one critical category each. Each condition included two learning groups. The stimuli were prepared in such a way that for one group of stimuli, [−unlearning], negative evidence and thus unlearning was not available, while for the other group, [+unlearning], negative evidence and thus unlearning was available. To achieve this, we manipulated the presentation order of the learning blocks. We will outline this by providing an analogy similar to experiment setups from the animal learning literature, for example, Kamin (1967, 1968)[1].

This difference in learning order is outlined in (1). The more abstract schema, where A = light, B = tone, X = food, and Y = toy is additionally shown in (2). The right arrow → indicates that an association is learned, and the stroked arrow ↛ that an association is unlearned. [±unlearning] groups within the actual learning experiment and simulation follow the same logic.

(1) [−unlearning]: (1) 💡 → 🍪 (2) 💡 🎵 → 🍪 🧸 (Test) 💡 → 🍪, {💡 → 🧸, 🎵 → 🧸}
[+unlearning]: (1) 💡 🎵 → 🍪 🧸 (2) 💡 → 🍪 (Test) 💡 → 🍪, {💡 ↛ 🧸, 🎵 → 🧸}

(2) [−unlearning]: (1) A → X (2) AB → XY (Test) A → X, {A → Y, B → Y}
[+unlearning]: (1) AB → XY (2) A → X (Test) A → X, {A ↛ Y, B → Y}

We will now explain the learning dynamics in more detail. Fig. 2 shows how the association between the cues for each outcome develops as a consequence of the learning order. The learning simulation is based on the Rescorla–Wagner equations (Rescorla & Wagner, 1972), which are described in more detail in Section 3.1.

The [−unlearning] groups first learn that a light cue on its own predicts food, and subsequently that both light and tone together predict food and toy outcome simultaneously. They cannot unlearn the association between the light cue and the toy outcome, because when they encounter the light together with food alone, they do not know yet that the toy is a possible outcome, so that the light–toy trials will not affect associations with the toy. To unlearn the light–toy association, learners would first have to expect the toy outcome, upon having experienced the light cue in the absence of the toy outcome. This kind of error is not available for the [−unlearning] groups.

The [+unlearning] groups first learn that a light cue, together with a tone cue, predicts both food and toy outcomes simultaneously. The groups subsequently learn that the light cue on its own predicts only the food. The first block would, therefore, lead to an equally strong association between light cue and food outcome as well as between light cue and toy outcome (the same holds for the tone–food and tone–toy association). The light–toy association is then

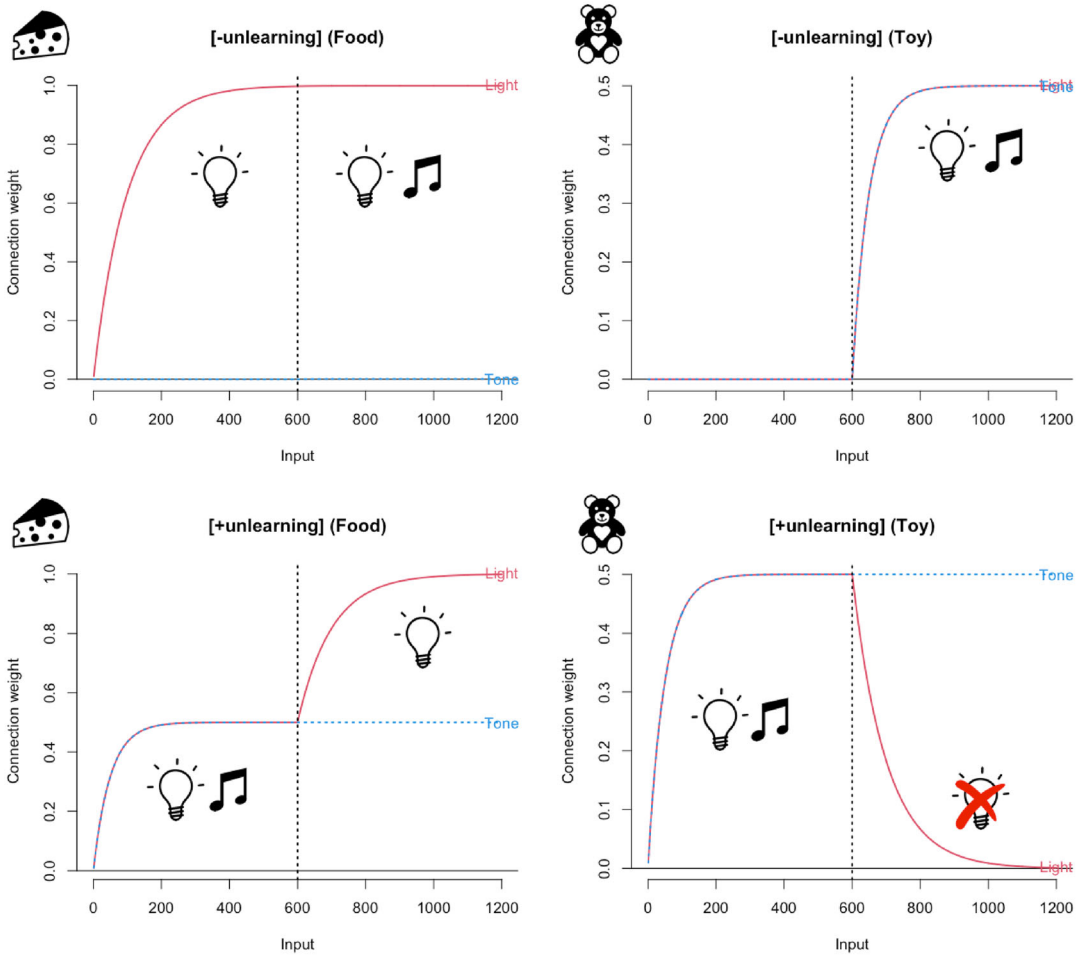

Fig. 2. Learning differences between [−unlearning] (top) and [+unlearning] (bottom). In [−unlearning], only light is associated with food, while both light and tone are associated with toy. The association between non-predictive light and toy is not weakened, because the light cue is initially encountered only with the food. Because of this, these trials cannot be used as negative evidence and the ambiguity is not resolved. In [+unlearning], the light–tone compound cue initially predicts the food–toy compound outcome. Afterwards, the light–cue alone predicts the food–outcome only. While light again has the strongest association with food, only tone is associated with toy: The light occurs in the absence of toy, so that this constitutes negative evidence, leading to a weakened light–toy association. Here, the ambiguity between cues is resolved. The simulation is based on the Rescorla–Wagner learning equations (see Section 3.1), using `edl` (van Rij & Hoppe, 2021).

unlearned in the second block, as the light cue now only predicts the food outcome. Crucially, this unlearning should, in turn, have consequences for the other cue–outcome association as well, namely that the tone cue on its own should end up as predictive of the toy outcome only, because the tone–toy association now has a stronger association weight than the light–toy one, in line with the comparator hypothesis (Miller & Matzel, 1988). We want to test whether

unlearning leads to the successful learning of both cue–outcome associations, compared to when no unlearning is available. In Section 3, we will explore how compound cues and outcomes influence the learning dynamics of the artificial learning experiment in more detail.

2.3. The learning groups

We created two learning groups for the two learning conditions. The DM_condition tested whether unlearning the *non-predictive* cue to the *diminutive* category, by associating it with the plural, would cause the *other, predictive* cue to be associated with the *diminutive* category instead. The PL_condition tested the consequence of unlearning for the plural category instead, that is, if unlearning the *non-predictive* cue to the *plural* category, by associating it with diminutive, would cause the *other, predictive* cue to be associated with this category instead. The names DM_condition and PL_condition thus indicate the category which we expect to be positively affected by the availability of unlearning. For example, in the case of the DM_condition, we wanted to know, if the word form [bef.nit͡ʃ] would be associated with the diminutive category. Likewise, for the PL_condition, we wanted to know if the word form [bef.ni] would be associated with the plural category. Importantly, participants in both conditions were not presented with distinct word forms that correspond to the critical category. Instead, they had to learn these associations based on exposure to the phonological cues within diminutive-plural word forms—word forms such as [bef.ni.t͡ʃi]. These two conditions allowed us to look into possible differences in unlearning between the [i] and [it͡ʃ] cue. Within a condition, both learning groups were presented with the exact same items. Only the learning block order differed.

For the DM_condition, the [−unlearning] group included as part of the first learning block word forms predictive of the singular [be.fan] and the plural [bef.ni] category. For the PL_condition, the diminutive category [bef.nit͡ʃ] replaced the plural category instead. For both conditions, the second learning block included different word forms predictive of the singulars again, but this time alternating with word forms predictive of the diminutive-plural category [bef.ni.t͡ʃi]. The first and the second learning blocks were reversed in the [+unlearning] groups, so that word forms predictive of singulars and diminutive-plurals occurred first, and different singulars, alternating with either plurals (DM_condition) or diminutives (PL_condition) followed second. The morphological categories were depicted as alien creatures, while the word forms were played auditorily. An overview of the learning groups and the learning order is shown in Table 2.

For the DM_condition, [−unlearning] had 32 singulars and 32 corresponding plurals as part of the first learning block; the second learning block had 32 different singulars and 32 corresponding diminutive-plural word forms. [+unlearning] reversed this order, so that the items of the second learning block in [−unlearning] were used for the first learning block in [+unlearning] instead, while items of the first learning block were used for the second one. For the test block, 16 diminutives and 16 plurals were used. The diminutives corresponded to the roots of the diminutive-plural word forms featured during learning. Sixteen novel plurals and corresponding diminutives were additionally used.

Table 2
Overview of the learning groups and the learning order. The form-meaning mappings are indicated by arrows, while the "Absent ✗"-column indicates the mapping that was not shown on its own. This mapping should be learned better when unlearning is available. All four groups were tested on [it͡ʃ] and [i] items only

Condition	Group	First Block	Second Block	Absent ✗ (But Expected)
DM_condition	[−unlearning]	[i]→ PL	[it͡ʃ]+[i] → DMPL	✗[it͡ʃ] → DM
	[+unlearning]	[it͡ʃ]+[i]→ DMPL	[i]→ PL	
PL_condition	[−unlearning]	[it͡ʃ]→ DM	[it͡ʃ]+[i]→ DMPL	✗[i]→ PL
	[+unlearning]	[it͡ʃ]+[i] → DMPL	[it͡ʃ]→ DM	

Abbreviations: DM, diminutive; DMPL, diminutive-plural; PL, plural.

For the PL_condition, [−unlearning] had 32 singulars and corresponding diminutives as part of the first learning block; the second learning block had 32 different singulars and 32 corresponding diminutive-plural word forms. [+unlearning] reversed this order, so that the items of the second learning block in [−unlearning] were used for the first learning block in [+unlearning] instead, while items of the first learning block were used for the second one. For the test block, 16 diminutives and 16 plurals were used. The plurals corresponded to the roots of the diminutive-plurals from learning. Sixteen novel plurals and corresponding diminutives were additionally used. The same roots occurred in both the DM_condition and PL_condition. While the singular and diminutive-plural word forms of the learning blocks and the novel word forms of the test were the same for all four groups, the roots for plurals in the DM_condition and the roots for diminutives in the PL_condition were shared across conditions. For example, [dep.ni] occurred as a plural word form in the DM_condition, while [dep.nit͡ʃ] occurred as a diminutive word form in the PL_condition. Word forms that were repeated in the test block—and which were based on diminutive-plurals—also had matching roots across conditions: For instance, where the DM_condition had [guk.lit͡ʃ], the PL_condition had [guk.li]—both based on the diminutive-plural word form [guk.li.t͡ʃi]. This ensured that the artificial language was comparable for all four groups.

To summarize, both conditions differed only in that the DM_condition included distinct plural word forms during learning, while the PL_condition included distinct diminutive word forms instead. The learning groups within the condition differed only in the presentation order of the learning blocks: In [−unlearning], singulars and either plurals or diminutives were presented first, with singulars and diminutive-plurals second. The order was reversed for the [+unlearning] groups. Within a condition, the exact same items were used for both groups.

3. Computational simulations

To understand how positive and negative evidence affect the cue–outcome associations of the artificial language, we modeled the learning process with `edl` (van Rij & Hoppe, 2021), as implemented in `R` (R Core Team, 2021). Cue–outcome associations within `edl` are updated using the Rescorla–Wagner equations (Rescorla & Wagner, 1972). Scripts

and data sets are made available at the following link: https://osf.io/9qt76/?view_only=ddcc3f8a30154b71801dcd17e8fbe030.

3.1. The Rescorla–Wagner learning equations

Connection weights are updated through a fully connected two-layer network, connected in a feed-forward manner. This means that for each subsequent learning event, the network updates the weights between cues to outcomes incrementally, on the basis of the formula shown in Eq. 1. Δw_{ij}'s value is calculated using Eq. 2 (Hoppe et al., 2022; Nixon, 2020; Rescorla & Wagner, 1972):

$$w_{ij}^{t+1} = w_{ij}^{t} + \Delta w_{ij}^{t} \tag{1}$$

$$\Delta w_{ij}^{t} = \begin{cases} \text{(a) } 0 & \text{if } \mathit{ABSENT}(c_i, t), \\ \text{(b) } \alpha_i\beta_1 \left(\lambda_j - \sum\limits_{[\text{Present}(c_i,t)]} w_{kj}\right) & \text{if } \mathit{PRESENT}(c_i, t) \,\&\, \mathit{PRESENT}(o_j, t), \\ \text{(c) } \alpha_i\beta_2 \left(0 - \sum\limits_{[\text{Present}(c_i,t)]} w_{kj}\right) & \text{if } \mathit{PRESENT}(c_i, t) \,\&\, \mathit{ABSENT}(o_j, t), \\ \text{(d) } 0 & \text{otherwise.} \end{cases} \tag{2}$$

The parameters λ_j, α_i and β_1/β_2 are adjustable: λ_j is a scaling parameter and stands for the total learnability of an outcome. α_i encodes cue salience, while β_1 reflects the salience of positive, and β_2 the salience of negative evidence. When a cue c_i is absent at learning trial t, no weight adjustments are made (a). The connection weights increase when a cue c_i and an outcome o_j are present in a learning trial t (b). This corresponds to positive evidence. If a cue c_i is present in a learning trial t, but an outcome o_j is absent, the connection weight decreases (c). This corresponds to negative evidence. In all other cases, no adjustments are made (d).[2] The magnitude of weight adjustments on a trial is a function of a prediction error, calculated by summing the connection weights of all present cues. This essentially reflects the error present during a learning trial: If the sum is high, that is, the presence or absence of an outcome is highly expected, the remaining error is small. As a result, only a small adjustment is made. For positive evidence, the sum is subtracted from λ_j (set to 1 as default) for negative evidence from 0, and then multiplied by the learning rate, either $\alpha_i\beta_1$ or $\alpha_i\beta_2$, (Baayen et al., 2011; Hoppe et al., 2022; Nixon, 2020; Rescorla & Wagner, 1972). `edl` uses the Rescorla–Wagner equations but collapses the $\alpha_i\beta_1$ and $\alpha_i\beta_2$ to a single parameter η set to 0.01. This corresponds to the Widrow–Hoff delta rule (Widrow & Hoff, 1960). The delta rule is able to model data using real-valued representations, while the Rescorla–Wagner equations only allow symbolic representations (Shafaei-Bajestan, Moradipour-Tari, Uhrig, & Baayen, 2023).

3.2. Coding and representation

We represented phonological word form as cues and morphological categories as outcomes. The phonological word forms were represented as bigrams (Baayen et al., 2011; Hoppe, van Rij, Hendriks, & Ramscar, 2020). Each word form was broken down into its constituent bigrams using the `ndl-package` (Arppe et al., 2018). We used bigrams to ensure that the phonological cues to the morphological categories—`plural`, `diminutive`, and `singular`—would emerge naturally as a consequence of learning. We also wanted to keep the amount of pre-specified representation units to a minimum, given that representation choices may affect model performance to a significant degree (Bröker & Ramscar, 2023).

The word forms were transcribed using the Serbo-Croatian orthography, which allowed us to represent each sound segment with one character. To better account for the relative position of a cue, word edges were additionally encoded as hash marks (Baayen et al., 2011; Saffran, Aslin, & Newport, 1996). For a word form such as `befnići`, all present phonological cues would simultaneously predict the morphological outcomes `diminutive` and `plural`.[3] The actual cues then consist of the bigrams `#b`, `be`, `ef`, `fn`, `ni`, `ić`, `ći`, and `i#`. In line with Hoppe et al. (2020, 2022), we included a constant background cue `**` which we can compare with the other cues. This allows us to better assess the relative informativity of the cues that we were interested in.

We would like to comment on two aspects of this representation: First, we needed to represent the diminutive-plural category by treating `diminutive` and `plural` as separate outcomes. This was necessary so that the models could learn something about the critical categories, that is, the ones for which no distinct word forms were presented (plurals for the PL_condition, diminutives for the DM_condition). Second, as each bigram cue was treated as a distinctive cue—with no information about the isolated characters (sounds) which make up these bigrams—our simulations will have no information about cue similarity available. For example, `i#` and `ić` both shared the same `i` character, but the difference between these cues was treated the same way as the difference between `i#` and `fn`—two cues which did not share any similar sound. This certainly is an oversimplification as human learners may also pay attention to the properties of the isolated sounds. Nevertheless, we deemed the bigram cues as more appropriate for two reasons: First, humans appear to not perceive each speech sound in an isolated form. This is not only due to practical reasons, as language processing would be slower, but also because information about certain sounds is only present in neighboring sounds (Port & Leary, 2005). Second, bigrams allow us to include information about the relative position of each cue: A word such as *socks* predicts a plural meaning only, because of the final [s] sound (corresponding to the `s#` bigram). Initial [s] or [s] sounds in general do not predict the plural meaning. Bigrams are, therefore, the smallest representation units that show both properties.

3.3. Simulation

We ran the models using the Rescorla–Wagner equations, as implemented in the `RWlearning` function in `edl` (van Rij & Hoppe, 2021). For each learning group, 128 learning trials occurred in total. The simulations are meant to model the learning experience

during the actual experiment. Thus, at the initial stage, the models which corresponded to the two [+unlearning] groups were first trained on the singular and diminutive-plural word forms only—64 trials in total. This constituted the first learning block. After this initial training, the networks were further trained with either new singular and plural word forms (DM_condition) or new singular and diminutive ones (PL_condition)—64 trials in total. This constituted the second learning block. For the two [−unlearning] groups, initial training was done with the second learning block instead, followed by the first. The trial order within the learning blocks was randomized. The learning network traced the association from each phonological (bigram) cue to each morphological (category) outcome. We assume that what is learned is on the one hand that phonological cues predict morphological outcomes to varying degrees, and on the other hand that each morphological outcome is predicted by a certain set of phonological cues. The cues for an outcome are then compared to each other with respect to how well each predicts this outcome. It is this comparison that affects participants' responses (Denniston et al., 2000; Miller & Matzel, 1988; Stout & Miller, 2007). For example, if cue A develops the same connection weight to both X and Y, but cue B develops a relatively higher weight to Y, and a relatively lower weight to X, the participants will more likely respond, upon exposure to A, with X, and upon exposure to B, with Y. Similar to Baayen et al. (2019), we, therefore, take the bigrams with the strongest connection weight for an outcome as the most likely cue to define this morphological category. The connection weights that develop for each cue–outcome association, therefore, reflect their relative strength and indicate if the association was learned.

3.4. Results

The final weight matrix for the two DM_condition groups contained 165 connection weights (55 cues × 3 outcomes). For the PL_condition groups, the final weight matrix contained 168 connection weights (56 cues × 3 outcomes). The `ć#` cue was additionally present in the PL_condition, as distinct diminutive word forms such as `befnić` occurred here. We asses how well the categories of the artificial language were learned, by inspecting the connection weights of the cues for each outcome separately, that is, on an outcome-by-outcome base. We assumed that learners would have learned that each cue predicts morphological categories to varying degrees but also that each morphological category is predicted by a certain set of cues. The cues for each category are thus compared to each other, in respect to how well they predict the category. This comparison then affects participants' responses. For example, if a cue A develops the same connection weight to both X and Y, but cue B develops a relatively higher weight to Y, and a relatively lower weight to X, then participants will more likely respond, upon exposure to A, with X, and upon exposure to B, with Y—in accordance with the comparator hypothesis (Denniston et al., 2000; Miller & Matzel, 1988; Stout & Miller, 2007). Similar to Baayen et al. (2019), we, therefore, take the bigrams with the strongest connection weight for an outcome as the most likely cue to define this morphological category. We will first present the results for the [−unlearning] and [+unlearning] groups of the DM_condition. The results are shown in Fig. 3.

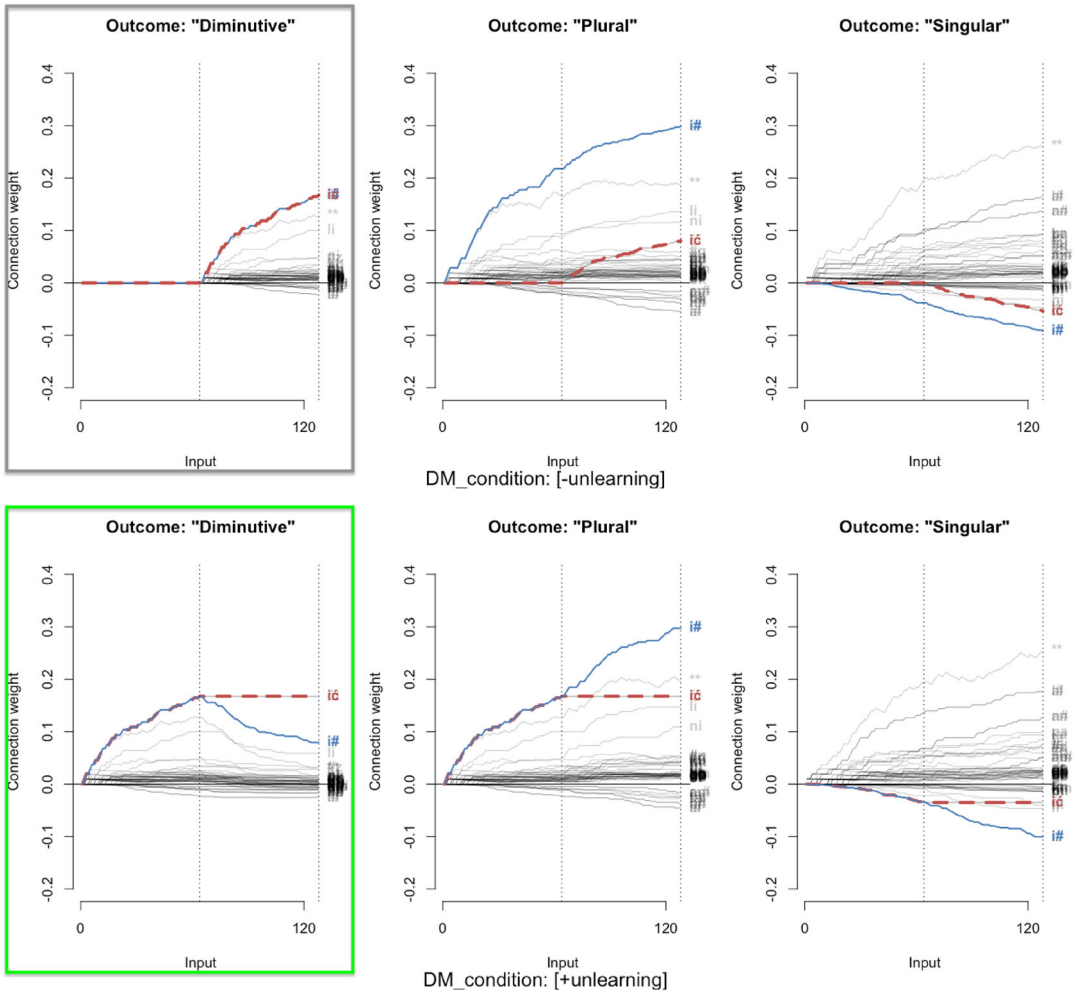

Fig. 3. Learned connection weights for bigram cues to each morphological outcome for [−unlearning] (top) and [+unlearning] (bottom) in the DM_condition. The first dashed line = the end of learning block I and the second dashed line = the end of learning block II. `i#` (in blue) is the predictive cue for the `plural`-outcome and `ić` (in red) the predictive cue for the `diminutive`-outcome. Highlighted simulations correspond to the critical category and show if unlearning occurs (green frame) or not (gray frame).

The results for the `diminutive` outcome show that unlearning only takes place in the [+unlearning] group. The connection weights between the non-predictive `i#` cue and the `diminutive` outcome decrease. During the first learning block, the model learns to associate both `i#` and `ić` cue with the `diminutive` outcome, as both together always predict the distinct outcomes `diminutive` and `plural`. During the second learning block, the model therefore expects a `diminutive` outcome, upon encountering the `i#` cue alone. However, as `i#` now predicts the `plural` outcome only, this results in prediction error, which causes the

i#-diminutive association to weaken. Ultimately, due to this weakening, the predictive ić cue is left with the strongest connection weight. In both groups, the ći cue also ends up with the same connection weight as ić.[4]

The results for the plural outcome show that the predictive i# cue develops the strongest connection weight to the plural outcome for both [−unlearning] and [+unlearning]. Interestingly, we also observe a blocking effect in the [−unlearning] group: The non-predictive ić cue only gains an insignificant amount of connection weight to the plural outcome as the association between i# and plural had already developed sufficient strength before the model encountered ić. This blocking effect is not present in [+unlearning], so that ić develops a stronger association in this group. However, its connection weight does not exceed the weight of the constant cue **, so that even for the [+unlearning] group the ić association is unlikely to interfere with the predictive i#-plural association.

Finally, the results for the singular outcome indicate that multiple cues develop a strong association with it, but that the constant ** cue actually ends up with the strongest weight. Bigrams in our case were apparently too coarse grained to capture the [a] vowel infix. Connection weights thus end up distributed across multiple cues—which in turn benefits the constant ** cue. While these cues were informative, they did not occur often enough to compete with the raw frequency of the constant cue. In respect to the unlearning effect though, both i# and ić develop the same negative connection weights in both groups, which means that both cues are not associated with the singular outcome.

The results for the [−unlearning] and [+unlearning] groups of the PL_condition are shown in Fig. 4. The model representations had the same design as in the DM_condition: The only difference was that the critical categories were exchanged—meaning that distinct word forms for the diminutive outcome, and not for the plural outcome, were presented. It should not come as a surprise then that the results for the PL_condition echo the previous ones: We observe the same difference between the [+unlearning] and [−unlearning] groups. For the diminutive outcome, ić, as well as ći, develop the strongest connection weights in both [+unlearning] and [−unlearning] groups. In the latter group, we again observed a blocking effect, which caused the other cues to develop weaker associations with the diminutive outcome. For the plural outcome, unlearning took place only in the [+unlearning] group, so that the ić cue decreased in connection weight, while in [−unlearning] both predictive i# and non-predictive ić cue end up with the same connection weight. For the singular outcome, i# and ić again develop a negative association with the constant ** cue developing the strongest connection weight. The exact values for the six bigrams with the highest connection strength for each category are included in the Supporting Information (see Tables S1– S4).

3.5. Discussion

The simulations confirm that unlearning will only take place when diminutive-plural word forms are encountered as part of the first block. When unlearning is possible, the actual bigram cues that correspond to the critical categories—diminutive for the DM_condition, and plural for the PL_condition—end up with the strongest connection weight, as the weights for the

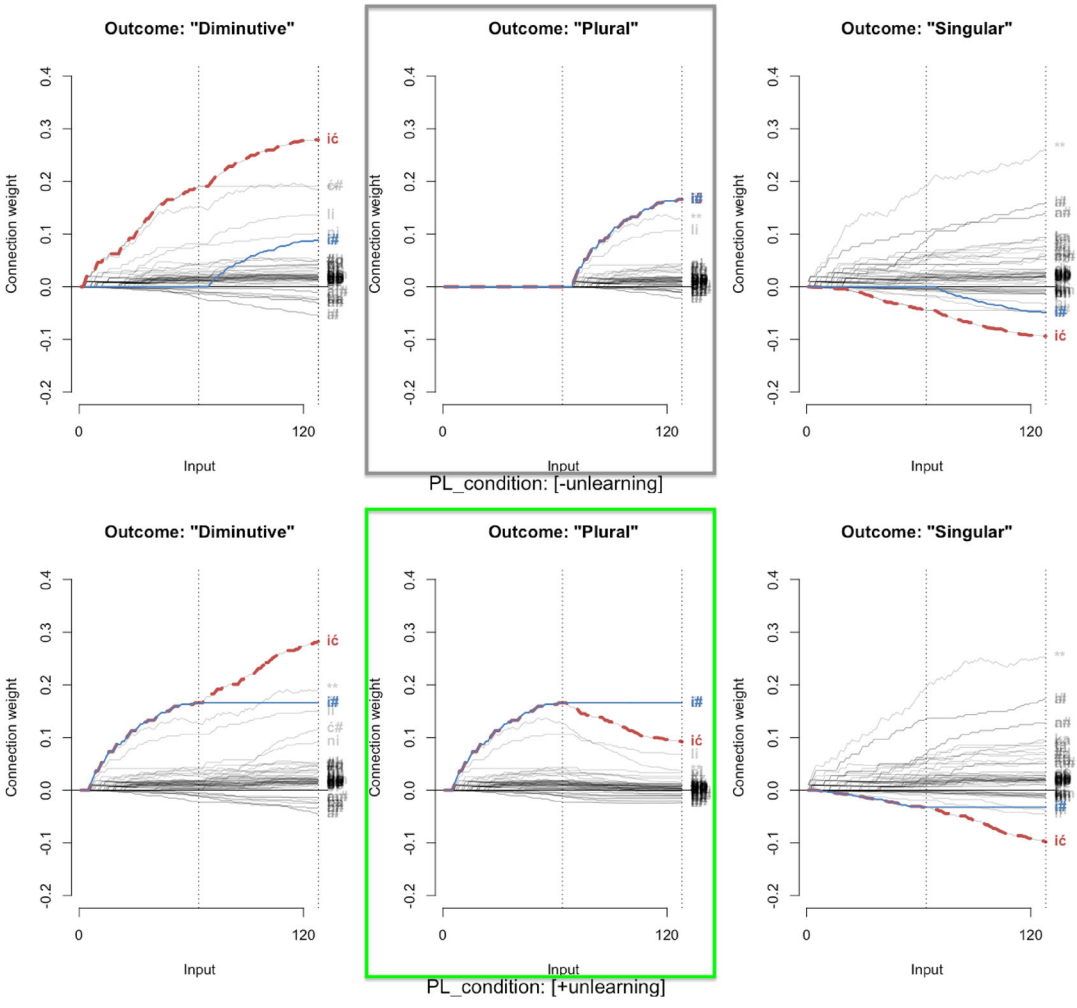

Fig. 4. Learned connection weights for bigram cues to each morphological outcome for [−unlearning] (top) and [+unlearning] (bottom) in the PL_condition. The first dashed line = the end of learning block I and the second dashed line = the end of learning block II. `i#` (in blue) is the predictive cue for the `plural`-outcome and `ić` (in red) the predictive cue for the `diminutive`-outcome. Highlighted simulations correspond to the critical category and show if unlearning occurs (green frame) or not (gray frame).

non-predictive cues decrease. When unlearning is not available though, predictive and non-predictive cue end up as the strongest predictor.

The results for the singular category show that the model did not learn one, but several cues to the category—though not these cues, but the constant cue ends up with the highest connection weight. This may raise the question of whether the singular category would be learned at all during the actual experiment. We believe that human learners would still manage to learn this category and that the simulation rather reflects the fact that the [a] infix is too

short. Regardless, we do not expect the learning of the singular category to interfere with the unlearning effect that we are actually interested in as the simulations for the singulars do not differ between groups.

With regard to the learning experiment, we believe that the unresolved competition between predictive and non-predictive cue in the [−unlearning] groups will be detrimental to the learning of the critical morphological categories. This ambiguity is only resolved for the [+unlearning] groups. We, therefore, anticipate that participants in the [+unlearning] groups will successfully learn that plurals are associated with the [i] cue, and diminutives with the [i͡ʃ] cue, while participants in the [−unlearning] group will only learn the association between [i] and plural (DM_condition) and between [i͡ʃ] and diminutive (PL_condition)—but not for the respective other, critical category.

Unlearning will only occur though if both positive and negative evidence affect learning. If negative evidence has no effect, we would not expect any difference between the [−unlearning] and [+unlearning] groups for the critical categories. The connection weight would not decrease, so that the ambiguity in both groups stays unresolved. If this were the case though, we would not observe that the critical category is better learned. Instead, the other category should be better learned—but not in the [+unlearning] groups, but instead in the [−unlearning] ones. This is due to the blocking effect being the only difference between groups. The competing cues for the unambiguous category would acquire a smaller connection weight in the [−unlearning] groups because they are encountered only after the predictive cue has been learned already. This means that for the DM_condition, plurals in the [−unlearning] group would be better learned than plurals in the [+unlearning] group, and for the PL_condition diminutives in the [−unlearning] group would be better learned than diminutives in the [+unlearning] group. Learning in this way is still based on the prediction error, although no negative evidence is taken into account. We will refer to such a model as the *positive-evidence only model* (Nixon, 2020).[5] In case both effects occur, however, we assume that unlearning has a stronger effect on learning than blocking. We believe the latter's influence to be negligible because the non-predictive cues end up with a substantially weaker connection strength—relative to the predictive ones—even when no blocking is present, as for the [+unlearning] groups.

To summarize, if both positive and negative evidence affect learning, unlearning will only take place in the [+unlearning] groups. This will then affect previously learned cue–outcome associations so that the ambiguity between two cues that predict two outcomes is resolved. We will only gather support for the role of negative evidence in morphophonological learning if the morphophonological patterns of the artificial language are indeed better learned for [+unlearning] than for [−unlearning].

4. Learning experiment

We programmed the learning experiment in PsychoPy (Peirce et al., 2019). We ran the experiment online in 2021, using Pavlovia (Bridges, Pitiot, MacAskill, & Peirce, 2020).

Table 3
Illustration of the visual stimuli of the alien species corresponding to the root /befn/

Singular (SG)	Plural (PL)	Diminutive (DM)	Diminutive-Plural (DMPL)
	 		 
[be.fan]	[bef.ni]	[bef.nit͡ʃ]	[bef.ni.t͡ʃi]

All scripts and data sets are available at https://osf.io/9qt76/?view_only=ddcc3f8a30154b71801dcd17e8fbe030.

4.1. Participants

One hundred twenty participants took part in the artificial language learning experiment. The participants were recruited through advertisements on social media platforms—the only prerequisite being sufficient proficiency in German.[6] Participants provided consent at the start of the experiment by pressing a key. They assigned themselves to one of the four groups based on their month of birth: For the DM_condition, January–March corresponded to [−unlearning] ($N = 32$, mean age = 27.06 years, range 19–52) and April–June to [+unlearning] ($N = 25$, mean age = 27.44 years, range 18–69). For the PL_condition, July–September corresponded to [−unlearning] ($N = 30$, mean age = 25.1 years, range 18–45), and October–December to [+unlearning] ($N = 29$, mean age = 28.1 years, range 19–67).

4.2. Material

We matched the word forms to visual depictions of singular, plural, diminutive, and diminutive-plural alien creatures. The pictures were taken from van de Vijver and Baer-Henney (2014). To depict the diminutive pictures, we reduced the size of the originals to 45%. Each species corresponds to a word root. In total, 81 different alien species occurred throughout the experiment—one of which was used only during familiarization. The four possible depictions for this species are shown in Table 3.

The first author, a male Serbo-Croatian native speaker with phonetic training, recorded the word forms—embedded in the carrier sentence *Forma _data ja* "The form _was provided." The items were realized with a falling pitch accent, so that they would reflect Serbo-Croatian word prosody (Lehiste & Ivić, 1986). The recordings were made in an anechoic chamber, using a Phantom Power 48V microphone, with a Sound Devices amplifier, a Marantz recorder, and a Transtec computer. The sampling rate was set to 48,000 Hz. We extracted the word

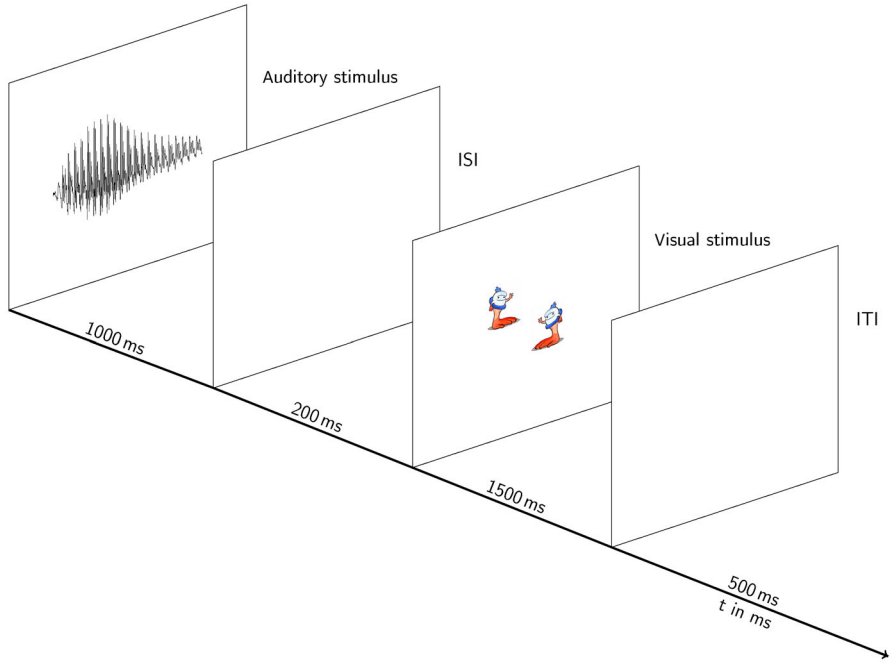

Fig. 5. Time course for a single trial (learning blocks I and II).

forms from the carrier sentence and scaled the intensity to 70 dB using a Praat script (Boersma & Weenink, 2021).

4.3. Procedure

The procedure was the same for all four learning groups. First, the four word forms that corresponded to the /befn/ root were played in random sequence, to familiarize the participants. Subsequently, the four visual depictions of the alien species were also played in random sequence. This ensured that at this stage, no mapping between form and meaning would be possible.

This familiarization was followed by the actual learning blocks. Which stimuli occurred in which learning block depended on the group and the condition (see Table 2 in Section 2.3). Within a trial, participants first listened to an auditory word form. The presentation time was set to 1,000 ms to ensure that each recording was played to full. After a short delay of 200 ms (the interstimulus interval), the visual depiction that corresponds to the word form was shown for 1,500 ms. The word forms thus served as cues in predicting the morphological categories (Nixon, 2020; Ramscar et al., 2010). After 500 ms (the intertrial interval), the experiment proceeded to the next trial. We randomized the trial sequence for each participant. Fig. 5 shows the time course for a single trial during the learning blocks.

Each learning block contained 64 trials (128 in total). In order for the participants to stay attentive, we included a pause screen between the first and the second learning blocks, which informed then that they had reached the halfway point.

The test block had a two-alternative forced choice design (2AFC). Participants had to decide, which depiction of the same alien species they believed to correspond to a given auditory word form by pressing one of two keys. The left key of the computer keyboard corresponded to the alien depiction on the left and the right key to the depiction on the right. After an input was registered, the experiment proceeded to the next trial with an intertrial interval of 500 ms. The experiment also proceeded to the next trial if participants took longer than 5,000 ms to answer. Participants did not receive any feedback about their choices throughout the experiment. We tested word forms that corresponded to the critical plural and diminutive categories because we were interested in the unlearning effect. We also tested the word forms of the category for which the appropriate form-meaning association was directly in order to assess whether participants had learned anything.

For half of the plural word forms, participants had to decide between plural and singular categories and for the other half between plural and diminutive. For half of the diminutive word forms, participants decided between diminutive and singular and for the other half between diminutive and plural. Diminutive-plurals were neither presented as word forms nor as possible answer. The correct answer side was counterbalanced. We did not test the other word forms because our error-driven learning simulations did not predict any difference between [±unlearning].

The test block consisted of 64 trials: 16 novel plurals and diminutives (both based on the same root) and 16 plurals and diminutives, repeated from the learning blocks. All trials were randomized. The time course of three trials within the test block is shown in Fig. 6. Incorrect choices (including timeouts) were coded as `0` and correct choices as `1`. The experiment took approximately 20 min to finish.

4.4. Prediction

We predict that the availability of unlearning will impact how well participants learn the category for which no distinct word forms were presented—that is, the critical category is only presented as part of the diminutive-plural configuration. Our error-driven learning simulations predict two effects resulting from contingency learning: Blocking in [−unlearning] and unlearning in [+unlearning] (see Section 3). It is unlikely that the weaker connection strength of the non-predictive cues in [−unlearning] compared to [+unlearning] would impact the learning of this category in any substantial way. The unlearning effect, in contrast, affects whether the ambiguity for the critical category is resolved or not. Our error-driven learning model, therefore, predicts that the [+unlearning] group in the DM_condition learns the diminutive category better, while the [+unlearning] group in the PL_condition learns the plural category better.

This only holds though if both positive and negative evidence affect learning. If negative evidence does not affect learning, the presence of the blocking effect in the [−unlearning] groups would lead us to expect that they have an advantage in learning the non-critical

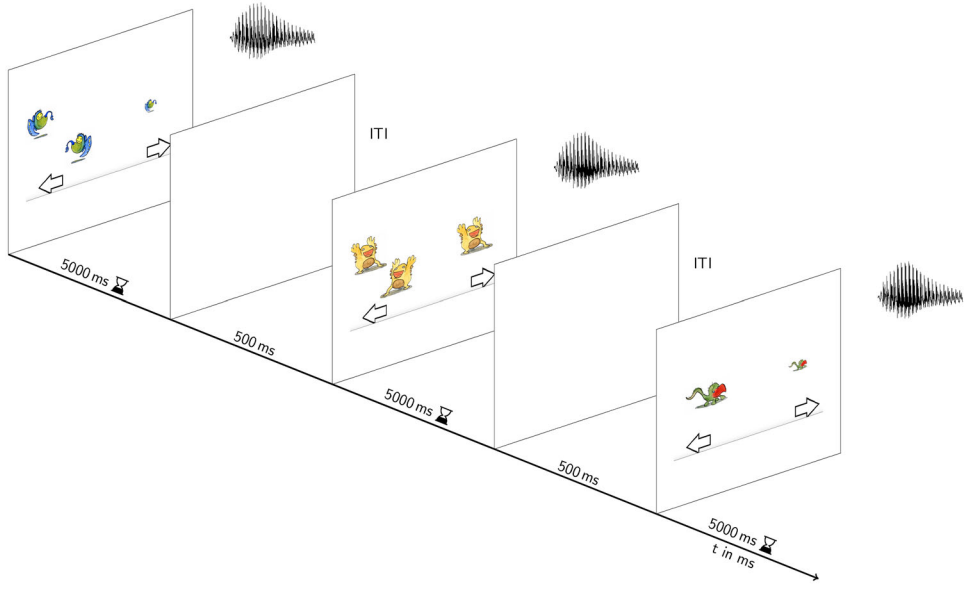

Fig. 6. Time course for three trials in the test block (choices between plural/diminutive, plural/singular, and diminutive/singular).

Table 4
Summary of the predictions for error-driven learning, positive-evidence only, and associative learning

	Frequency	Blocking	Unlearn	Prediction
Error-driven learning	No	Yes	Yes	[−unlearning] < **[+unlearning]**
Positive-evidence only	No	Yes	No	**[−unlearning]** > [+unlearning]
Associative learning	Yes	No	No	**[−unlearning]** = **[+unlearning]**

category (plural in the DM_condition and diminutive in the PL_condition) instead. This positive-evidence only model would thus predict that the [−unlearning] order leads to better learning than the [+unlearning] order. As a final prediction, we will consider whether learning is not based on contingency but on contiguity instead. Learning then is based solely on the frequency of the present cue and present outcome (Adriaans & Kager, 2010; Ambridge et al., 2015; Baer-Henney & van de Vijver, 2012). In this case, no difference between the [−unlearning] and [+unlearning] groups should emerge as the same input is present for both groups. We will refer to this frequency-based learning as associative learning—subsuming both frequency-based statistical learning (Baer-Henney et al., 2015; Bybee, 2010) and distributional learning (Maye et al., 2002). The crucial difference between these three predictions is summarized in Table 4.

We take the proportions of correct choices as reflective of how well a category is learned. The result for the critical category of each condition will indicate whether unlearning the non-predictive cue to an outcome also has an effect on the successful learning of the predictive

cue to the same outcome. If this is the case, the critical morphological category will be better learned in the [+unlearning] groups. In turn, the results for the other category will act as a reference to how good the participants were able to learn the unambiguous category during the experiment. Here, an advantage for the learning of the non-critical category in [−unlearning] would mean that only the blocking effect was present while unlearning either has no effect on the corresponding outcome choice or simply did not occur.

4.5. Results

One hundred twenty participants took part in the experiment. We removed two participants because they timed out in each trial and, therefore, had a total accuracy of 0. We removed another two because they always chose the incorrect answer option—which we believe indicates that they somehow misunderstood the experiment task. The first two were in the [+unlearning] group of the DM_condition, and the other two were in the [+unlearning] group of the PL_condition. After removing these, we ended up with 7,424 data points (116 participants × 64 trials). Within the DM_condition, 2,048 data points corresponded to the [−unlearning] group and 1,600 to the [+unlearning] group. Within the PL_condition, 1,920 data points corresponded to the [−unlearning] group and 1,856 to the [+unlearning] group. Participants did not provide an answer in only 91 cases.[7] We analyzed the data using the packages `tidyverse` (Wickham et al., 2019), `lmerTest` (Kuznetsova, Brockhoff, & Christensen, 2017), `ggplot2` (Wickham, 2016), and `visreg` (Breheny & Burchett, 2017) within the `RStudio` environment (RStudio Team, 2021) in `R` (RStudio Team, 2021).

We will report the results for each condition separately. The correct choice accuracy for diminutive and plural word forms for the [−unlearning] and [+unlearning] groups in the DM_condition is shown in Fig. 7. The violin plots show the distribution of mean values for each participant.[8] For the DM_condition, participants in the [−unlearning] group had an overall accuracy of 0.68 ($SD = 0.466$) and 0.69 ($SD = 0.462$) for diminutive and plural each. Similarly, participants in the [+unlearning] group had an overall accuracy of 0.70 ($SD = 0.459$) for diminutive but 0.85 ($SD = 0.360$) for plural. Participants in the [+unlearning] group appear to have learned to associate word forms for the non-critical plural category better than participants in the [−unlearning] group with no difference between groups in the critical diminutive category. We fitted a generalized linear mixed effects model for the accuracy of data in the DM_condition. The model summary is shown in Table 5.

As fixed effects we entered `Group` (two levels: `[−unlearning]`, `[+unlearning]`), `Category` (two levels: `diminutive`, `plural`) and `Sequence`—the trial order during the test as well as their interactions (pairwise and three way). We included `Sequence` because test trial order has been shown to affect learning (Heitmeier, Chuang, & Baayen, 2023). A change in accuracy in later trials may reflect either participants' change in confidence or a reevaluation of their cue–outcome mappings. For random effects, we entered the by-Word slope for the effect of `Group` and the by-Participant slope for the effect of `Category`.[9] We replaced the by-Word slope with the intercept for `Word` as the former led to singular fit (R syntax: `accuracy ~ Group * Category * Sequence + (1| Word) + (1+Category | Participant), family = binomial (link = ''logit''))`. The intercept of the model was set to the `Diminutive` category of the [−unlearning] group (`Sequence` = 31.5).

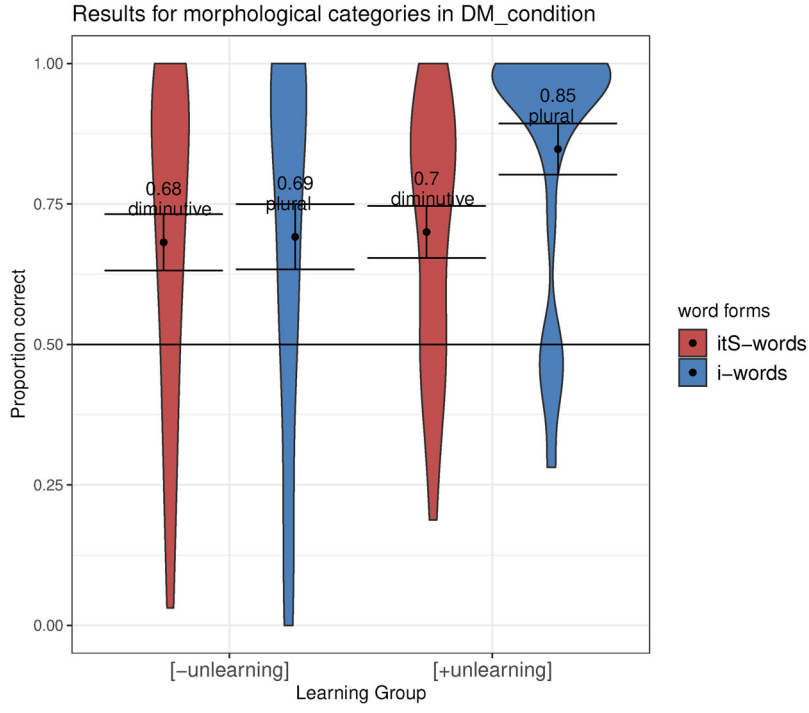

Fig. 7. Correct choice proportions for diminutive and plural word forms in the DM_condition. Error bars show the 95% confidence interval.

While the fixed effects `Group`, `Category`, and `Sequence` were all not significant, the two-way interactions and three-way interaction were all significant (`Group` × `Category`: Est. = 3.065, $SE = 0.844$, $z = 3.633$, $p < .001$; `Group` × `Sequence`: Est = 0.0213, SE = 0.007, $z = 3.111$, $p < .01$; `Category` × `Sequence`: Est. = 0.029, $SE = 0.007$, $z = 4.036$, $p < .001$; `Group` × `Category` × `Sequence`: Est. = −0.051, $SE = 0.011$ $z = -4.727$, $p < .001$). The plotted summaries of the model are included in the Appendix (see Fig. A.1). Compared to the [−unlearning] group, participants in the [+unlearning] group had higher accuracy for the plural category and did also improve in general. However, as indicated by the negative coefficient of the three-way interaction, participants in this group did not improve as much for plurals (compared to diminutives) as the [−unlearning] group (see also Fig. B.1 in Appendix B). Neither `Group`, `Category`, nor `Sequence` on their own had an effect. This likely indicates that participants in the [−unlearning] group improved in accuracy for plurals during the test, but not for diminutives, while the [+unlearning] group in fact improved for both categories. The lower accuracy for plurals in [−unlearning] in comparison to [+unlearning] may follow from the difference in association strength after learning. The association between [i]-final plural word forms and the plural category was stronger in the [+unlearning] group, while the association between [it͡ʃ] final diminutive word forms and the diminutive category did not differ between groups. What the three-way interaction likely indicates is that due to this

Table 5
Summary of the generalized linear mixed effects model for accuracy in the DM_condition—[−unlearning] and [+unlearning]. The model is dummy coded

Random Effects Groups	Name	Variance	Std. Dev.	Corr
Word	(Intercept)	0.196	0.443	
Participant	(Intercept)	2.758	1.661	
Participant (Category = Plural)	(Slope)	7.224	2.688	–0.290
Number of observations: 3,648				
Groups: Word, 64; Participant, 57				

Fixed effects	Estimate	Std. Error	z-value	Pr(>|z|)	
(Intercept)	1.070	0.352	3.040	<.01	**
Group = [+unlearning]	−0.580	0.509	−1.140	.254	
Category = Plural	−0.349	0.561	−0.622	.534	
Sequence	0.004	0.005	0.829	.407	
Group: Category	3.065	0.844	3.633	<.001	***
Group: Sequence	0.021	0.007	3.111	<.01	**
Category: Sequence	0.029	0.007	4.036	<.001	***
Group: Category: Sequence	−0.051	0.011	−4.727	<.001	***

$p \leq .001$ ***; $p \leq 0.01$ **; $p \leq 0.05$ *; $p \leq 0.1$.

learning order difference, the [−unlearning] group was initially less certain about the [i]-plural association. They became more confident in this association during later trials possibly because some trials required them to choose between diminutive and plural for [i]-final words, and some trials required them to choose between singular or diminutive for [it͡ʃ]-final words. They thus continued learning. In [+unlearning], both categories got better but the certainty for plurals after training was already high enough. These results show that learning order had an influence on learning. However, unlearning the [i]-diminutive association apparently did not benefit the diminutive but the plural category instead. This finding is unexpected in regard to the predictions of full error-driven learning, where we would have expected for the [+unlearning] group to learn diminutives and not plurals better. Interestingly, these results are also not predicted by positive-evidence only models, which predict better performance for plurals in [−unlearning], nor associative learning, which predicts no difference whatsoever.

Next, we assess how well the morphological categories were learned in the PL_condition. The correct choice accuracy for diminutive and plural word forms in the [−unlearning] and [+unlearning] groups is shown in Fig. 8. The violin plots show the distribution of mean values for each participant. For the PL_condition, participants in the [−unlearning] group had an overall accuracy of 0.78 ($SD = 0.418$) for diminutive, with only 0.42 ($SD = 0.495$) overall accuracy for plural. Participants apparently had great difficulties learning the critical plural category. Participants in [+unlearning] instead have comparable results for both diminutive and plural, showing an overall accuracy of 0.78 ($SD = 0.414$) for diminutive and 0.73 ($SD = 0.447$) for plural. Participants apparently learned the critical plural category in [+unlearning] but not in [−unlearning].

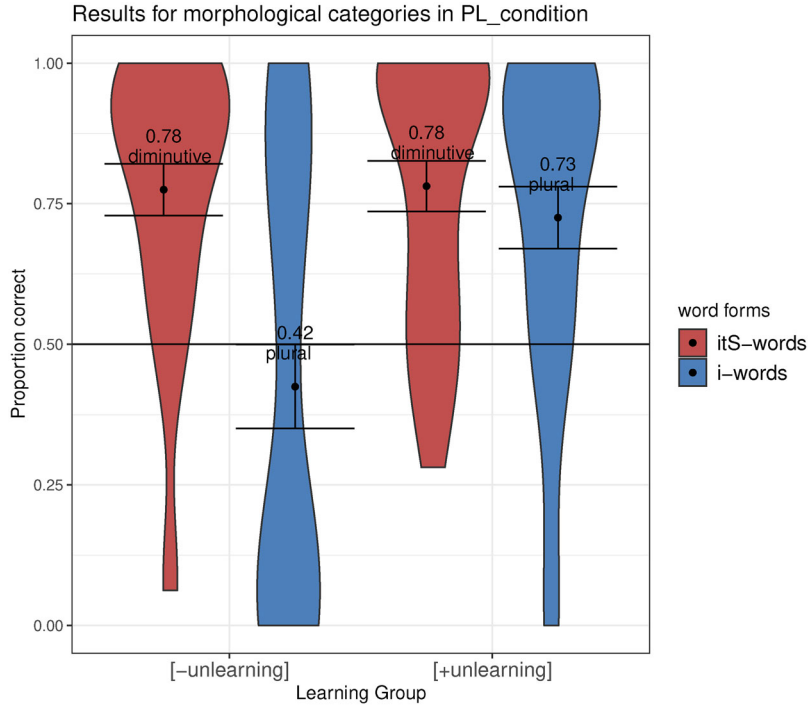

Fig. 8. Correct choice proportions for diminutive and plural word forms in the PL_condition. Error bars show the 95% confidence interval.

We fitted a generalized linear mixed effects model for the accuracy of data in the PL_condition. The model summary is shown in Table 6. For the PL_condition, we entered `Group` (two levels: [−unlearning], [+unlearning]), `Category` (two levels: `diminutive`, `plural`), and `sequence` as fixed effects as well as their interactions (pairwise and three way). As random effects, we again entered the by-Word slope for the effect of `Group` and the by-Participant slope for the effect of `Category` but had to replace the by-Word slope with the intercept for `Word`, as including it led to a singular fit (R syntax for the final model: `accuracy ~ Group * Category * Sequence + (1| Word) + (1+Category | Participant), family = binomial (link = ''logit''))`. The intercept of the model was set to the `Diminutive` category of the [−unlearning] group (`Sequence = 31.5`).

The fixed effects `Category` and `Sequence` were not significant, but `Category` ended up significant (Est. $= -3.550$, $SE = 0.649$, $z = -5.474$, $p < .001$). Plurals were learned worse than diminutives. The two-way interaction between `Group` × `Category` was significant (Est. $= 1.987$, $SE = 0.903$, $z = 2.200$, $p < .05$) as well as the two-way interaction between `Category x Sequence` (Est. $= 0.018$, $SE = 0.008$, $z = 2.337$, $p < .05$). No other interactions were significant (although `Group` × `Sequence` approached significance). The plotted summaries of the model are included in the Appendix (see Fig. A.2). The model shows that the accuracy for plural was worse than for diminutive in both [−unlearning] and [+unlearning]

Table 6
Summary of the generalized linear mixed effects model for accuracy in the PL_condition, [−unlearning] and [+unlearning]). The model is dummy coded

Random Effects Groups	Name	Variance	Std. Dev.	Corr
Word	(Intercept)	0.182	0.426	
Participant	(Intercept)	4.943	2.223	
Participant (Category = Plural)	(Slope)	8.910	2.985	−0.370
Number of observations: 3,776				
Groups: Word, 64; Participant, 59				

Fixed effects

	Estimate	Std. Error	z-value	Pr(>|z|)	
(Intercept)	1.912	0.464	4.125	<.001	***
Group = [+unlearning]	0.848	0.661	1.284	.199	
Category = Plural	−3.550	0.649	−5.474	<.001	***
Sequence	0.005	0.005	1.023	.306	
Group: Category	1.987	0.903	2.200	<.05	*
Group: Sequence	−0.014	0.007	−1.890	.059	.
Category: Sequence	0.018	0.008	2.337	<.05	*
Group: Category: Sequence	0.005	0.011	0.462	.644	

$p \leq .001$ ***; $p \leq 0.01$ **; $p \leq 0.05$ *; $p \leq 0.1$.

(significant effect of category), but that this effect actually stems from the worse performance for the plurals in the [−unlearning] group only as indicated by the significant interaction between group and category. The participants in the [+unlearning] group, therefore, learned both diminutive and plural categories equally well. This effect was still present even when controlling for the interaction between group, category, and sequence. While the significant interaction between category and sequence indicated that plurals ratings got better with later test trials, this effect was uniform for both groups, given that the three-way interaction was not significant (see also Fig. B.2 in Appendix B). This is different from the DM_condition, where plurals for the [+unlearning] group did not improve during later trials as much as for the [−unlearning] group. This indicates that the difference between [−unlearning] and [+unlearning] stems from the difference in learning order and the availability of unlearning. Fully in line with the predictions of error-driven learning, the participants in [+unlearning] learned the plural category better than participants in [−unlearning] as they could unlearn the association between [i͡ʃ] and plural, which resulted in a stronger relative association between [i] and plural. Furthermore, accuracy results for plural word forms in [−unlearning] were especially bad with only 0.42 correct choices, indicating that most participants in this group were not able to learn the plural category.

4.6. Discussion

We tested whether unlearning a cue–outcome association would also affect other cue–outcome associations. We hypothesized that an ambiguous morphological category is better

learned if the cue for the other, unambiguous category is encountered second. When this cue is encountered with one outcome, while the other is absent, the connection weight between the present cue and the absent outcome is reduced. This reduction in turn allows the absent, predictive cue for the absent outcome to end up with a higher relative weight—which reflects the successful learning of this cue–outcome association.

We structured the input so that unlearning was only available for one group. We did this by manipulating the learning order. The absence of one outcome in the presence of a cue would constitute negative evidence and trigger unlearning only in the [+unlearning] group. For the other group, this absence is not registered as the absence, given that participants have not encountered the outcome at this point. Depending on the condition, unlearning was predicted to allow either the diminutive or the plural category to be better learned. Our results for the PL_condition confirm this as the plural category was better learned in the [+unlearning] group. In the DM_condition, the [+unlearning] group also seems to have a learning advantage but surprisingly also for the plural category—and not for the diminutive category, which was predicted to benefit from unlearning.

With regard to our predictions, neither positive-evidence only models nor (contiguity-based) associative learning are able to explain this result. While the positive-evidence only model does indeed predict better learning for the non-critical category, it does so not for the [+unlearning] group—as observed here—but for the [−unlearning] group instead. It predicts this due to the blocking effect. Associative learning in turn is not able to explain this result as it does not predict any difference between groups. Generally speaking, our results appear to be in line with the predictions of error-driven learning models: For both conditions, the availability of negative evidence—and thus unlearning—actually improved the learning results—although not to the same degree.

Why is it that we observe an asymmetry between the conditions? We believe that this asymmetry results from a difference in perceptual salience between the phonological cues that were unlearned: The critical diminutive category in the DM_condition may have been easier to learn, because [it͡ʃ]—the predictive cue for the diminutive category—was more salient than [i]—the predictive cue for the plural category. Even though no explicit evidence for the [it͡ʃ]-diminutive association was present, [i] did not compete as much with [it͡ʃ] due to the difference in salience. There are both general phonetic and language-specific phonological reasons that could lead to greater salience of [it͡ʃ] compared to [i]. For instance, both [it͡ʃ] and [i] share the same vowel, with the former cue containing an additional consonant—an affricate sibilant. Affricates are complex segments that involve a complete closure of the vocal tract—followed by an abrupt and sharp frication noise (Ladefoged & Johnson, 2014). The [ʃ] portion in turn involves a high amount of energy in higher frequency regions (Zsiga, 2013). Such higher energy may be especially salient after an initial closure phase. Due to these properties, [t͡ʃ] may contrast with other consonant sounds that involve more evenly distributed acoustic energy in the artificial language. A phonological reason relates to the informativity (or functional load) of /t͡ʃ/ in German. As this sound occurs rather infrequently (Wiese, 1996), its actual occurrence could more easily be noticed. Less frequent stimuli are often considered

more salient than high-frequency stimuli, possibly due to their higher informativity (Boswijk & Coler, 2020). It is likely that these factors contributed to the greater salience of [it͡ʃ].

In order to test if controlling for the salience of [it͡ʃ] in `edl` simulations would predict the asymmetry between conditions, we reran the simulations, controlling for perceptual salience. We did this using a modified version of the Rescorla–Wagner function from Smolek and Kapatsinski (2023). This function allowed us to adjust the α-parameter for each bigram cue beforehand (see Eq. 2). To reflect the higher salience of the [t͡ʃ] sound, we set the α-parameter to 0.15 for all bigram cues that contained the sound (i.e., `ić, ći, ć#`). For all other cues, the α-parameter was set to the default value of 0.1. The simulation was otherwise run with functions from the `edl`-package (van Rij & Hoppe, 2021). The results of this simulation are plotted in Fig. 9 for the DM_condition and in Fig. 10 for the PL_condition. Both groups in the DM_condition learn that the [it͡ʃ] cue is most predictive of the diminutive category, while only the [+unlearning] group in the PL_condition learns that the word final [i] cue is most predictive of the plural category. For the [−unlearning] group in the PL_condition, [it͡ʃ] is most predictive of both diminutive and plural categories, so that the plural category is not learned. These simulations show that controlling for the [it͡ʃ] cue's greater salience results in a simulation more in line with our empirical findings.

We would like to acknowledge though that opting for an explanation grounded in perceptual salience itself may be problematic. For instance, it is not exactly clear, how to best determine perceptual salient in general (see MacLeod, 2015). More importantly, though, we may run into pitfalls of confusing cause and effect: Was one cue better learned, because it is more salient, or do we assume that the cue is more salient because it was better learned? (see Boswijk & Coler, 2020 for a discussion on salience). Still, for the above-mentioned reasons, we feel justified in assuming a possible role of salience here.

During learning, the more salient [it͡ʃ] cue overshadowed the less salient [i] cue. Although this is not reflected by our simulations, this overshadowing effect could additionally explain why we not only observe stable results for diminutives in the DM_condition but also worse results for plurals in the [−unlearning] group. The [it͡ʃ]-plurals association may interfere with the [i]-plural association because the [it͡ʃ] + [i] → diminutive-plural trials occurred second here, which resulted in a more recent [it͡ʃ] → plural exposure. This could have caused learners to question whether [i] still predicted plurals, which would explain the worse plural accuracy at the beginning of the test. By reevaluating the relative weights for diminutive and plural outcomes during the test, participants in [−unlearning] were able to resolve this confusion. Both cue salience and cue competition appear to interact in complex ways. While not all aspects of this interaction are clear, it is the case that a slight difference in learning order can make negative evidence available, and thus result in distinct learning outcomes.

Because salience interacts with cue competition, the consequences are especially severe for the PL_condition. When no unlearning is available, the more salient [it͡ʃ] overshadows the less salient [i] cue, resulting in the predictive [i]-plural association not being learned at all. Participants in the [−unlearning] group did not learn the plural category because they likely ended up associating the [it͡ʃ] cue with both the diminutive (predictive) and plural (non-predictive)

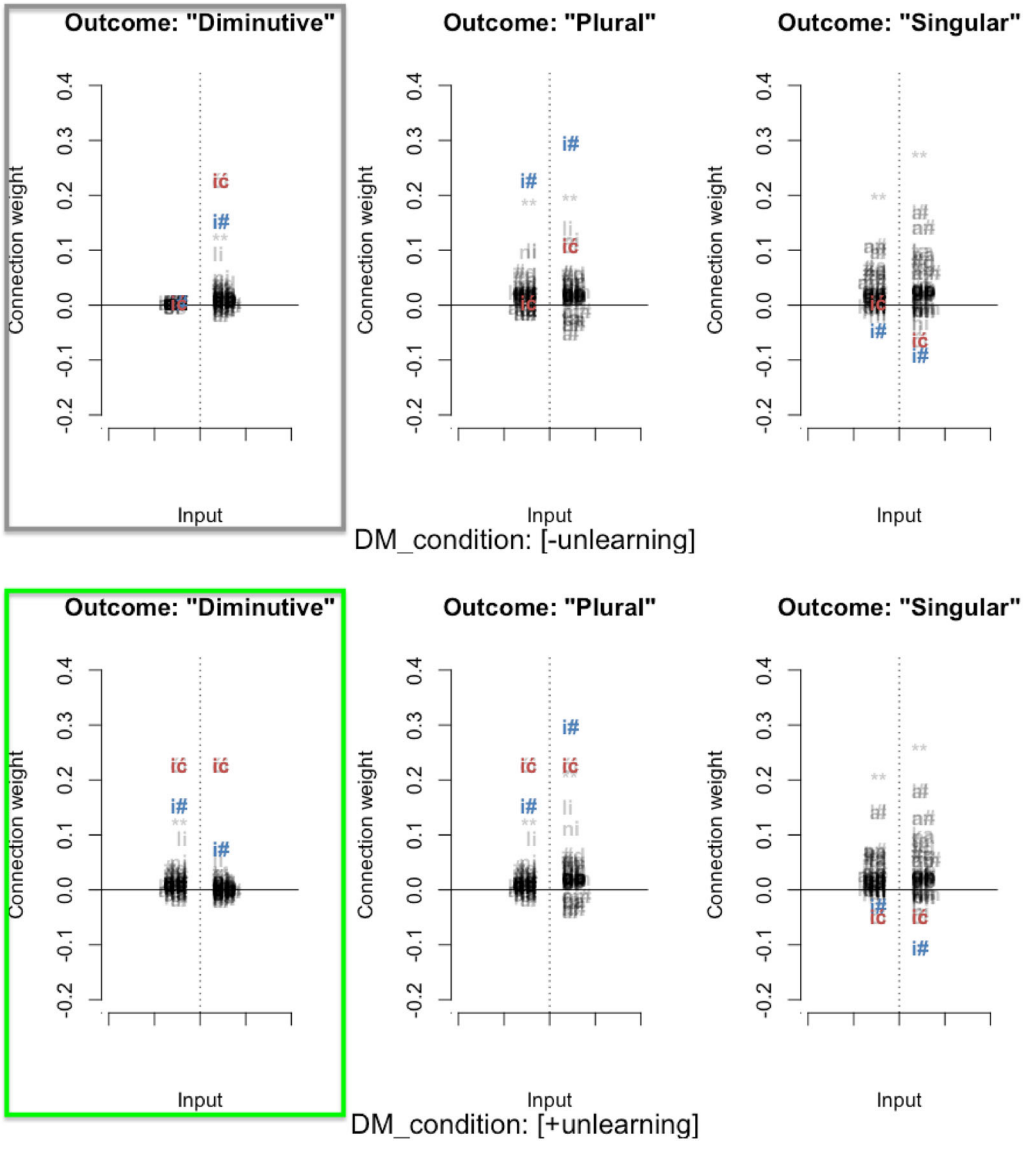

Fig. 9. Total connection weight for bigram cues to each morphological outcome for [−unlearning] (top) and [+unlearning] (bottom) in the DM_condition. The α-parameter for `ić`, `ći`, and `ć#` was set to 0.15 to reflect the higher salience of these cues. The dashed line divides learning blocks I and II. Highlighted simulations correspond to the critical category and show if unlearning occurs (green frame) or not (gray frame).

category. This is reflected in the simulations in Fig. 10 as `ić` has the strongest connection weight for both diminutive and plural in the [−unlearning] group. Given that participants had unambiguous evidence for the [it͡ʃ]-diminutive association, participants were more likely to pick the diminutive category when deciding between plural and diminutive categories for

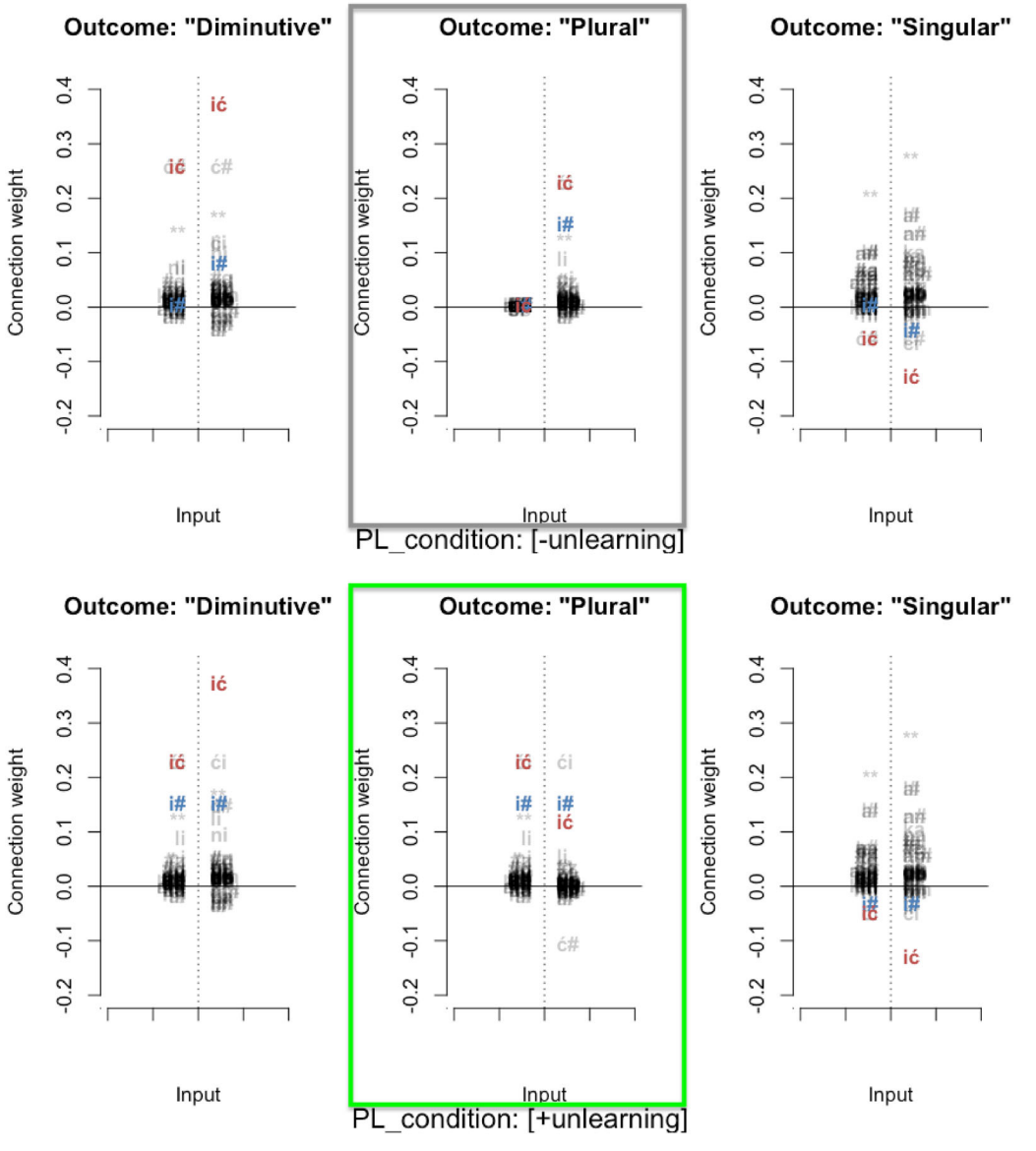

Fig. 10. Total connection weight for bigram cues to each morphological outcome for [−unlearning] (top) and [+unlearning] (bottom) in the PL_condition. The α-parameter for `ić`, `ći`, and `ć#` was set to 0.15 to reflect the higher salience of these cues. The dashed line divides learning blocks I and II. Highlighted simulations correspond to the critical category and show if unlearning occurs (green frame) or not (gray frame).

[it͡ʃ] word forms, as reflected by the higher connection weight for ić-diminutive compared to ić-plural. This category was therefore learned. When unlearning was possible though, the non-predictive association between [it͡ʃ] and plural weakened, which in turn allowed the predictive association with [i] to emerge. This is reflected by the i# cue having the strongest connection weight out of all cues for the plural outcome. Under this account, both groups in the DM_condition learn the [it͡ʃ]-diminutive and the [i]-plural association, while only the [−unlearning] groups in the PL_condition learn the [it͡ʃ]-diminutive association. These results are in line with error-driven learning. Our results show that negative evidence affects not only the cue–outcome association that is unlearned but also other cue–outcome associations (Kapatsinski, 2018a; Nixon, 2020; Ramscar et al., 2010, 2013).

Our results indicate that learners can make use of negative evidence, which helps them learn the corresponding form-meaning mappings, even when faced with ambiguity. This effect is especially important in case where non-discriminative—but more salient cues—would otherwise overshadow discriminative cues of lesser salience. During learning, more salient stimuli are more quickly picked up on as they attract attention. This attention generates strong expectations about a cue's relative importance in predicting future outcomes, that is, the cue's predictive value. If a more salient cue is present, while an expected outcome fails to occur, this results in a higher amount of prediction error (Nixon & Tomaschek, 2023). From this, a stronger unlearning effect follows.

5. General discussion

We started out with the question of what kind of evidence learners use to learn language. In linguistic theory, which predominantly considers learning theories that rely on contiguity learning (Baer-Henney & van de Vijver, 2012; Baer-Henney et al., 2015; Bybee, 2010; Maye et al., 2002), it is often assumed that only positive evidence affects learning. Error-driven learning, in contrast, offers a differentiated view of the evidence used by learners (Harmon et al., 2019; Nixon, 2020; Nixon & Tomaschek, 2021); learners use both positive and negative evidence to the extent that they help minimize prediction errors. There is a good deal of evidence showing that learners rely on both positive and negative evidence (Harmon et al., 2019; Nixon, 2020; Nixon & Tomaschek, 2021); (Ramscar, Dye, & Klein, 2013), but one gap in our knowledge concerns the role of negative evidence when faced with ambiguity: What happens when at first two cues—say AB—predict two outcomes—say XY—and in the next step, one of the cues—for example, A—predicts only the X outcome: Does the presence of A in the absence of Y result in their association being unlearned, and, as a consequence, B and Y being learned instead? If this is the case, it would constitute evidence for learning from negative evidence. Error-driven learning theory (Hoppe et al., 2022; Nixon, 2020; Ramscar et al., 2010; Rescorla & Wagner, 1972) predicts this because it allows for negative evidence to influence learning. We found evidence for such unlearning in morphophonological learning.

To test for this effect, we taught German-speaking adults novel morphophonological patterns in an artificial language learning experiment. We constructed the language so that

unlearning was available for one group but not for the other. One group first learned word forms that expressed two morphological categories, that is, diminutive-plural. At this stage, it was not clear which part of the word predicts the diminutive meaning and which part of the word predicts the plural meaning. Later, participants were presented with word forms that expressed only one morphological category—for example, diminutive. This, we hypothesized, would not only lead to an increase in the weight of the cue for the actual diminutive outcome but also to a decrease in the weight between this cue and the absent plural outcome. Importantly, this decrease in weight results in the absent cue being associated with the absent outcome instead as this association ends up with the highest relative weight. The plural category would then also be learned. If the learning order is reversed though, this unlearning effect is not available. In this case, the absent cue will not be associated with the absent outcome. We tested this prediction in one condition with the absent outcome being the diminutive category (DM_condition)—for which a more salient cue was predictive of the category—and in one condition with the absent outcome being the plural category (PL_condition)—for which a less salient cue was predictive.

Our results are in line with the predictions from error-driven learning, as we did observe better accuracy when unlearning was available for both conditions. For the PL_condition, participants better learned to associate the absent cue with the absent plural category, with no difference between groups for the diminutive category. For the DM_condition, unlearning did not cause participants to better learn that the absent cue is associated with the absent diminutive category but instead led to a stronger association between plural cue and plural category. Here, we observed no difference for the diminutive category, which was actually predicted to be learned better. The unlearning effect we observed differed depending on the properties of the unlearned phonological cues. Unlearning thus took place and affected learning, simply on the basis of negative evidence.

5.1. Asymmetry due to salience differences

We observed an asymmetry between conditions, which is likely due to a difference in perceptual salience between the phonological cues for plurals and diminutives. A difference in perceptual salience is able to explain why the diminutive category in the DM_condition was equally well learned in both groups, while the plural category was better learned in the [+unlearning] group instead: For the DM_condition, the ambiguity between cues is resolved through the more salient cue overshadowing the less salient one. This explains why the diminutive category was learned, even when no unlearning was available. When the ambiguous word forms are encountered more recently though—as was the case in the [−unlearning] group—the more salient cue of the diminutive category also interferes with the association of the cue that was predictive for the plural category, which resulted in plural word forms being classified worse. Overshadowing thus benefited learning the [it͡ʃ]-diminutive association in the DM_condition, while it was detrimental for learning the [i]-plural association in the PL_condition. It was beneficial in the former because it allowed the category to be learned despite the ambiguity. For the latter, it was detrimental because it caused the more salient [it͡ʃ] cue to also be associated with the plural category. The association between the

non-predictive—but salient—cue and the plural category was only weakened when this cue was encountered while the plural category was absent. This suggests that unlearning is especially important in cases where more salient cues would otherwise overshadow less salient ones.

Salience effects are not unheard of within the error-driven learning literature. Differences in cue salience, for example, may affect the speed of learning, as association weights for more salient cues develop faster (Kapatsinski, 2023b; Miller, Barnet, & Grahame, 1995). Salience itself is also influenced by temporal order as temporally closer cues may also be more salient (Smolek & Kapatsinski, 2023). While such effects can be accounted for by adjusting the learning rate per cue, other salience effects are harder to deal with. Miller et al. (1995), for example, point out that the Rescorla–Wagner learning equations have difficulties in dealing with stimulus salience—especially when considering the interaction with the blocking effect. Heckler, Kaminski, and Sloutsky (2006) show that blocking effects are attenuated if the additional cue is more salient than a pre-trained blocker cue. Miller et al. (1995) argue that more salient additional cues may even cause backward overshadowing as a pre-trained cue could completely lose its association with an outcome (Miller et al., 1995). We may observe a similar effect for the plural category in the [−unlearning] group. It could be the case that the more salient cue interfered with the less salient one even though participants had already encountered unambiguous word forms for this category.

While salience effects can be accounted for by adjusting the α and β parameters of the Rescorla–Wagner equations (see Mackintosh, 1975), this solution is ideal as it introduces additional researcher degrees of freedom. Hoppe et al. (2022), for example, generally advise against hard coding salience in error-driven learning models, in contrast to Rescorla and Wagner (1972). They argue that salience effects should be captured by the model as a consequence of previous learning. In other cases, models that explicitly try to capture salience effects should be used instead. Salience effects can be captured implicitly by using recent error-driven learning models that implement incremental learning with real speech input. Shafaei-Bajestan et al. (2023), for example, trained a linear discriminative model (Baayen et al., 2019) on MEL-scaled frequency bands in order to simulate the human auditory perception of isolated words. Because Mel-scaled frequency bands mimic the functions of the human cochlea, information about salience is also included. In this way, hard coding a different learning rate for more salient cues can be avoided. Encoding phonological cues as continuous acoustic cues generally appears to be a promising direction for future studies.

Previous studies that have investigated the learning mechanisms in language learning also observe effects of perceptual salience: For instance, Olejarczuk et al. (2018) found that phonetic category representations of a rise–fall tone shifted towards the rarer, more surprising tokens, but overall ratings favored tokens with greater pitch excursions—independent of the distribution's skew. This could be interpreted as large pitch excursions being more likely to stand out, and thus to develop a stronger association to an outcome. Nixon (2020) attested to the blocking effect for the learning of a novel phonetic category with English speakers. Yet, she observed the effect only when pre-training was done with a lexical high tone as cue—but not when pre-training was done with nasalized vowels. If cues differ in salience, so that high tones are more salient than nasalized vowels, this effect follows naturally. As we did

not investigate the interaction between learning mechanisms and cue salience in this study explicitly, future studies have to investigate such effects in a more systematic way.

Another aspect that might have had an effect on our results comes from the learners' knowledge of German. Because German has syncretic diminutive and diminutive-plural forms, for example, *Hündchen* [hʏntçən] can refer to either one small, or multiple small puppies. As learners in the DM_condition encountered separate word forms for diminutive-plurals and plurals, they may have transferred their learned mapping from word forms corresponding to the diminutive-plural category to the diminutive category as well. Harmon and Kapatsinski (2017) found that adult English learners associated an infrequent phonological cue that corresponds to a singular diminutive cue with a general diminutive meaning, so that they used this cue to express both diminutive-singular and diminutive-plural meaning. Because diminutive (singular) word forms in our study were quite similar to diminutive-plural word forms (e.g., [bef.nit͡ʃ] vs. [bef.ni.t͡ʃi]), this mapping would then result in the diminutive category also being learned. While this explanation is able to account for why the diminutive category was learned equally well in both groups, it does not address why we observe a difference in learning for the plural categories instead. If word forms for diminutive-plural are simply mapped to the diminutive category, this mapping would be available for both learning groups. Perceptual salience appears to us as a more likely explanation because it is able to account for the difference in plurals as well as for the asymmetry between conditions.

5.2. Negative evidence in language learning

While previous studies also found unlearning effects when both positive evidence for the predictive, and negative evidence for the non-predictive association were present (Nixon, 2020; Ramscar et al., 2010), it was unclear, if negative evidence also affects learning in cases where learners are not certain about the informativity of other cues. Our results suggest that even in these cases, negative evidence is used, as unlearning occurred and allowed to resolve ambiguity between the cues for an absent outcome. While Harmon et al. (2019) observed unlearning in phonetic category learning only when other discriminative cues to the same outcome were explicitly shown and usable, our results indicate that, for morphophonology, unlearning also occurs during trials where no additional cues occur. Learners can resolve ambiguities of the kind present in our study by unlearning a previous cue–outcome relationship. Without negative evidence, learning appears to be less efficient, as certain ambiguities cannot be resolved—as has been the case for the plural category in the PL_condition.

In general, our results show that learners not only make use of positive (Baer-Henney & van de Vijver, 2012; Baer-Henney et al., 2015; Bybee, 2010; Maye et al., 2002) but also use negative evidence (Harmon et al., 2019; Nixon, 2020; Rescorla & Wagner, 1972; Ramscar et al., 2010). Error-driven learning accounts for the role of negative evidence by allowing associations between present cues and absent outcomes to be unlearning. Allowing negative evidence to affect language learning has considerable implications for research on the learnability of language and the poverty-of-stimulus problem (Bowerman, 1988; Elman, 1993; Marcus, 1993; Pearl, 2022; Pullum & Scholz, 2002; Ramscar et al., 2013; Rohde & Plaut, 1999; Saxton, 2000; Scholz & Pullum, 2002; Tomasello, 2005). Ramscar et al. (2013), for

instance, argue that the learnability problem does not arise if children can correct themselves through negative evidence. However, the notion of negative evidence has been met with suspicion (Bowerman, 1988; MacWhinney, 2004; Marcus, 1993; Yang, 2016), so that alternative proposals argue for positive evidence only models (see Pearl, 2022, for an overview).

When viewed from an error-driven learning perspective, negative evidence instead appears neither too complex nor too unconstrained to be used in language learning. Within error-driven learning, ungrammatical associations are only unlearned if learners do not encounter the structures they expect. Negative evidence may have been previously regarded as unreliable, due to the conception that some form of corrective feedback is needed. While negative evidence from one's own prediction error may incorporate other sources of negative evidence—such as corrective feedback, recasts, or requests for clarification (Bowerman, 1988; Saxton, 2000; Scholz & Pullum, 2002)—it is neither the only source of negative evidence nor necessary for learning. Both learning and unlearning centers on the learners own prediction and their prediction error (Nixon, 2020; Olejarczuk et al., 2018; Ramscar et al., 2010, 2013; Rescorla & Wagner, 1972).

If negative evidence affects learning, a question that follows is whether tracking the occurrence and non-occurrence of events is cognitively plausible. It appears quite unlikely that association weights for all absent outcomes are adjusted, given that during a learning event, there is only a limited amount of outcomes that are present, while a vast amount of outcomes is absent (Hollis, 2019). There are always more absent than present stimuli. In our setup for the [−unlearning] groups, when a phonological cue occurs with a specific morphological outcome, learners cannot make use of negative evidence for other outcomes, if they do not know yet that other outcomes will eventually occur. However, learners know from their experience that expressions for small or multiple creatures exist, so in principle could also consider these as outcomes. But what non-occurrences does the learner actually track? Hollis (2019) suggests that a cognitively more plausible model involves updating association weights only for outcomes which have been encountered, so that weights for absent outcomes are not directly decreased. This can be achieved by using the Widrow–Hoff learning rule (Widrow & Hoff, 1960), using gradient descent (Rumelhart, Hinton, & Williams, 1986), or the vector learning model (Hollis, 2019). The vector learning model makes similar predictions as the Rescorla–Wagner model but represents cues and outcomes as vectors in an *n*-dimensional space. A cue vector moves closer to an outcome vector if they co-occur together. From this, it follows that present cues get dissociated from absent outcomes—unlearning based on negative evidence—as the cue vectors also move further away from absent outcome vectors. This manages to both fix the problem computationally and to constrain the outcome space. Human learners may additionally consider all encountered stimuli as possible outcomes—at least temporally. Infants are overwhelmed by the "blooming buzzing confusion" (James, 1890) as they cannot discern which contingencies are probable. Later though, they are able to reduce their space of expectations, possibly due to other mechanisms, for example, generalization or formation of a conceptual space (Goldberg, 2019). Eventually, learners are able to discern which (expected) outcomes are relevant and which are not. An outcome's absence, in the presence of a cue, will then influence their associations—whether in a direct or indirect way. How learners decide which cue–outcome pairings are relevant and which are not is difficult to answer though.

Although error-driven learning is able to incorporate negative evidence in a cognitively plausible way, it is not the only learning theory that allows for negative evidence. One such theory is *Bayesian learning*. Regier and Gahl (2004) observed that Bayesian learning is able to incorporate negative evidence due to its sensitivity to the absence of inputs, similarly to the way, adult learners in Ramscar et al. (2013) were sensitive to the absence of *wug*. Regier and Gahl (2004) conclude that Bayesian learning is able to account for implicit negative evidence and that other mechanisms, for instance, innate knowledge, are not needed. Bayesian learning is also able to account for the vast space of possible non-occurrences, by proposing that negative evidence is less informative than positive evidence (McKenzie & Mikkelsen, 2007). However, Bayesian learning models are based on a perfect and rational learner. This assumption does not always hold (Ramscar et al., 2013). Bayesian learning may possibly explain the unlearning effect we found, though it is unclear, whether it would also predict other learning-order-related effects (Baayen, Shaoul, Willits, & Ramscar, 2016b; Kruschke, 2006). A comparison with Bayesian learning falls beyond the scope of the present study though.

5.3. Implications for natural language learning

In our study, the availability of negative evidence was followed by a difference in learning order. Our results, therefore, imply that the way the input is structured by the learner also matters during learning. For instance, Elman (1993) suggests that learners profit from starting with a simpler input, which gradually increases in complexity—in a way which resembles the changes in the brain's storage capacity during maturation. In contrast, Rohde and Plaut (1999) found better learning when learners are exposed to the full complexity of the language first before the complexity gradually decreases. Our results are more in line with the latter findings as our learners learned the artificial language better when exposed to the configural diminutive-plural word forms first and to the single category forms second. These results are also in line with Arnon and Ramscar (2012), who observe that both nouns and their articles are better learned, if learners are first exposed to the full article+noun sequence and then to the noun alone. It is often assumed that during language development, learners transition from unanalyzed, holistic chunks to analyzed, specific parts (Arnon, 2021). Such transition could be attributable to domain general error-driven learning mechanisms as a general consequence of learning to better discriminate relevant cue dimensions. In order to achieve this, negative evidence appears to be crucial.

It is unclear if error-driven learning mechanisms—and thus negative evidence—can be used by infants right from the start. Nevertheless, the way the input is structured appears to have very real consequences for learning. Ramscar et al. (2013) have shown that children learn to associate object names based on informativity, that is, how well the object predicts the name. If objects A and B predict the label *dax*, and objects B and C predict *pid*, then the novel label *wug* is equally well associated with both A and C—but not with B, as B occurs more often in the absence of *wug* than the other two. Adults instead base their association on logical inference, associating *wug* with B. However, the adult learners in our study also based their associations on informativity—not on logical inference. If our learners were driven by logical inference, both the [−unlearning] and [+unlearning] groups would have learned

the morphological categories equally well, given that both groups could deduce what the predictive cue for the absent outcome may be (if A → X, and AB → XY, then B → Y). This was not the case. Whether these learning mechanisms are available right from the beginning and used during all stages of life is beyond the scope of this study. The learning behavior of the adult learners in our study is accounted for by error-driven learning models that use the Rescorla–Wagner equations (Rescorla & Wagner, 1972).

We have shown that it is possible for unlearning to occur even in the absence of explicit positive evidence. As we have conducted an artificial language learning experiment, this included learning in a rather artificial setting—involving only a limited number of cues and outcomes. However, this raises the question of whether the representations that we used realistically reflect natural language learning. Advances in natural language processing allow the use of large language models to capture many natural language phenomena. While semantic outcomes are often discrete (Kapatsinski, 2023b), recent studies have moved away from discrete representations. Nieder et al. (2023), for example, use continuous cue–outcome representations to model Maltese inflection, while Heitmeier et al. (2023) successfully modeled trial-by-trial effects of a lexical decision experiment in the same way. These studies use one-hot encoded vectors to represent phonology, and word embeddings to represent meaning, replacing the Rescorla–Wagner equations with the similar, but computationally more powerful Widrow–Hoff delta rule (Widrow & Hoff, 1960). In the context of our study, MEL-frequency bands (Shafaei-Bajestan et al., 2023) could be used to represent the auditory stimuli and visual embeddings (Shahmohammadi, Heitmeier, Shafaei-Bajestan, Lensch, & Baayen, 2024) to represent the visual stimuli. We expect that the unlearning effect we have observed will also emerge with such continuous representations. Using more realistic representations has the advantage of accounting for both phonological similarity through continuous acoustic cues, and meaning similarity through visual embeddings. Furthermore, the perceptual salience effect is also likely to emerge from the values within the MEL-frequency bands or the durational differences between [it͡ʃ] and [i]. While we have opted for discrete representations of cues and outcomes for the sake of simplicity because we were only interested in the error-driven learning mechanism of unlearning, a more realistic representation such as outlined above should be considered for future studies.

A related question is whether the effect we have observed is also present in more complex learning scenarios. In natural language learning settings, a vast number of cues compete with many more outcomes. For example, the [ɐ] sound alone expresses 60 different morphological functions in German (Tomaschek & Ramscar, 2022). Likewise, in German, multiple affixes can express plural morphology (Heitmeier, Chuang, & Baayen, 2021). Additionally, morphological categories, such as plural, are also rich in meaning, showing systematic variation, depending on the semantic class of the noun (Nikolaev, Chuang, & Baayen, 2022; Shafaei-Bajestan, Uhrig, & Baayen, 2022). These intricacies make it hard to understand the precise role unlearning may have in arriving at these representations. While it is possible that the absence of an expected outcome is more easily noticed under experimental conditions, it should still be the case that learners rely on the same learning mechanisms even in more complex situations. A child may initially assume that all four legged animals are DOGS but eventually discovers that some four legged animals are CATS instead. The cue *four legged*

to DOGS needs to be weakened, so that cues that better predict DOGS are learned (e.g., *has a certain body size and head shape; does bark*). A child may also assume that the biological sex of a person matches the grammatical gender of its noun, though for many cases this does not hold, for example, the German noun *Mädchen* "girl" takes the neuter gender (Samuel, Cole, & Eacott, 2019). Using the biological sex to predict the grammatical gender thus needs to be unlearned. This shows that both phonological and semantic cues together, in addition to various other modalities, need to be considered as well. Moving away from simple, discrete representations and considering multimodal cue/outcome structures, in addition to a more differentiated representation for form and meaning will allow us to better approximate how learners make use of both positive and negative evidence during language learning. Error-driven learning is one approach to language learning that is compatible with life-long changes in a person's language. As suggested by individual differences in adults (Dąbrowska, 2015; Ramscar, 2022; Ramscar & Port, 2016), speakers do not generalize to a fixed grammar, which is shared by the speaker community, but they approximate a finite state. Likewise, cue–outcome associations are not static but are continuously adjusted on the basis of error. This allows for (language) learning to continue throughout life.

The present study provided empirical evidence for implicit negative evidence to affect morphophonological learning. We have shown how the effect of negative evidence follows from domain general learning mechanisms, captured by error-driven learning models. This kind of negative evidence affects learning only though, if a learner entertains a hypothesis, or expects an outcome first, and later fails to observe the necessary input needed for the hypothesis/prediction to hold. Learners thus make use of both the available and the unavailable evidence, which allows them to learn language in an efficient way.

6. Conclusion

We investigated how unlearning a cue–outcome association affects the learning of other, previously learned associations. Our results show that encountering a cue, when an expected outcome is absent, leads to unlearning. This unlearning takes place even when no further positive evidence for other cue–outcome associations is present. By having this kind of negative evidence available, ambiguous cues can be learned. This learning mechanism predicts that attending to greater chunks first, and more specific ones later, leads to better learning, as learners are then able to learn language not only from present cues and outcomes but also from present cues and absent outcomes. Importantly, both learning and unlearning do not proceed in isolation, but instead affect the entire system, so that the cue which best predicts the outcome is learned. The mechanism which is responsible for this is the same error-driven learning mechanism available in other domains, such as motor skill or knowledge learning. Through unlearning previous associations that prove to be non-predictive, that is, non-informative, novel associations are formed. Learning associations between present cues and present outcomes, while simultaneously unlearning the ones between present cues and absent outcomes allows for life-long and efficient language learning—despite impoverished or skewed input.

Acknowledgments

We would like to thank the audience of the EDLL 2022 as well as all attendees of an online discussion meeting held at the University of Tübingen in 2022 (in particular: Titus von der Malsburg, Elnaz Shafaei-Bayestan, Jessie Nixon, Motoki Saito) for their useful comments. We also like to thank our two reviewers, Harald Baayen and Jacolien van Rij, as well as our editors, Padraic Monaghan and Rick Dale, who all provided useful comments on this paper and helped us to improve it.

Open access funding enabled and organized by Projekt DEAL.

Open Research Badges


This article has earned Open Data and Open Materials badges. Data and materials are available at https://osf.io/9qt76/.

Endnotes

1 The actual stimuli (cues and outcomes) for the language-learning experiment are explained in Section 2.3.

2 The Rescorla–Wagner equations do not include learning about absent cues, that is, a cue being absent, while an outcome is present. Learning thus proceeds from cues to outcomes in a unidirectional manner. (See Van Hamme & Wasserman, 1994 and Nixon, Poelstra, and van Rij, 2022 for a discussion on negative cue evidence.)

3 Given that alien pictures were used as outcomes during the actual learning experiment, each learning trial would also contain information about the alien species. To account for this, we included the lexical root as additional outcome, for example, `befn`. Including the roots as outcomes did not affect our results in any way, so that we decided to only report the simulations with the morphological categories as outcomes.

4 We will only focus on the relative importance of the `ić` cue though, as we did not test word forms ending in [it͡ʃi]—that is, diminutive-plurals—in the actual learning experiment.

5 The results for the simulation with only positive evidence (parameter β$_2$ set to 0) are included in the Supporting Information (see Figs. S3 and S4).

6 Most participants were either L1 German native speakers ($N = 90$) or bilingual L1 German speakers with an additional L1 ($N = 19$). We initially removed the non-L1 German speakers from the analysis, but as removing them did not change our results we decided to keep them in the final analysis.

7 Excluding these participants had no effect on our statistical analysis.

8 We have aggregated the results of 2AFC contrasts in [it͡ʃ] and [i] word forms. The separate results for each 2AFC contrast (dm/pl, dm/sg, pl/dm, pl/sg) can be found in the Supporting Information (see Figs. S1 and S2).

9 To control for the correct answer location, we also tried out separate models in which we included either the random intercept for `correctAnswerSide`, that is, whether right or left was correct or the random intercept for `response`, that is, whether participants

choose left or right. Including either measures led to a singular fit, due to the estimates being zero or near zero, so that our final model contained neither. This holds for both the DM_condition and PL_condition.

References

Adriaans, F., & Kager, R. (2010). Adding generalization to statistical learning: The induction of phonotactics from continuous speech. *Journal of Memory and Language*, *62*(3), 311–331.

Albright, A., & Hayes, B. (2011). Learning and learnability in phonology. In *The handbook of phonological theory* (pp. 661–690). Chichester, England: Blackwell.

Ambridge, B., Kidd, E., Rowland, C. F., & Theakston, A. L. (2015). The ubiquity of frequency effects in first language acquisition. *Journal of Child Language*, *42*(2), 239–273.

Arnon, I. (2021). The starting big approach to language learning. *Journal of Child Language*, *48*(5), 937–958.

Arnon, I., & Ramscar, M. (2012). Granularity and the acquisition of grammatical gender: How order-of-acquisition affects what gets learned. *Cognition*, *122*(3), 292–305.

Arppe, A., Hendrix, P., Milin, P., Baayen, R. H., Sering, T., & Shaoul, C. (2018). *ndl: Naive discriminative learning*. R package version 0.2.18. https://CRAN.R-project.org/package=ndl

Baayen, R. H., Chuang, Y.-Y., & Blevins, J. P. (2018). Inflectional morphology with linear mappings. *The Mental Lexicon*, *13*(2), 230–268.

Baayen, R. H., Chuang, Y. Y., Shafaei-Bajestan, E., & Blevins, J. P. (2019). The discriminative lexicon: A unified computational model for the lexicon and lexical processing in comprehension and production grounded not in (de)composition but in linear discriminative learning. *Complexity*, *2019*, 1–39.

Baayen, R. H., Hendrix, P., & Ramscar, M. (2013). Sidestepping the combinatorial explosion: An explanation of n-gram frequency effects based on naive discriminative learning. *Language and Speech*, *56*(3), 329–347.

Baayen, R. H., Milin, P., Đurđević, D. F., Hendrix, P., & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review*, *118*(3), 438.

Baayen, R. H., Shaoul, C., Willits, J., & Ramscar, M. (2016a). Comprehension without segmentation: A proof of concept with naive discriminative learning. *Language, Cognition and Neuroscience*, *31*(1), 106–128.

Baayen, R. H., Shaoul, C., Willits, J., & Ramscar, M. (2016b). Comprehension without segmentation: A proof of concept with naive discriminative learning. *Language, Cognition and Neuroscience*, *31*(1), 106–128.

Baer-Henney, D., Kügler, F., & van de Vijver, R. (2015). The interaction of language-specific and universal factors during the acquisition of morphophonemic alternations with exceptions. *Cognitive Science*, *39*(7), 1537–1569.

Baer-Henney, D., & van de Vijver, R. (2012). On the role of substance, locality, and amount of exposure in the acquisition of morphophonemic alternations. *Laboratory Phonology*, *3*(2), 221–249.

Blaser, R., Couvillon, P., & Bitterman, M. (2004). Backward blocking in honeybees. *Quarterly Journal of Experimental Psychology Section B*, *57*(4), 349–360.

Boersma, P., & Weenink, D. (2021). *Praat: Doing phonetics by computer* [Computer program]. Version 6.1.50. http://www.praat.org/

Boswijk, V., & Coler, M. (2020). What is salience? *Open Linguistics*, *6*(1), 713–722.

Bowerman, M. (1988). The 'no negative evidence' problem: How do children avoid constructing an overly general grammar? In *Explaining language universals* (pp. 73–101). Oxford, England: Basil Blackwell.

Breheny, P., & Burchett, W. (2017). Visualization of regression models using visreg. *The R Journal*, *9*(2), 56–71.

Bridges, D., Pitiot, A., MacAskill, M. R., & Peirce, J. W. (2020). The timing mega-study: Comparing a range of experiment generators, both lab-based and online. *PeerJ*, *8*, e9414.

Bröker, F., & Ramscar, M. (2023). Representing absence of evidence: Why algorithms and representations matter in models of language and cognition. *Language, Cognition and Neuroscience*, *38*(4), 597–620.

Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes*, *10*(5), 425–455.

Bybee, J. (2010). *Language, usage and cognition*. Cambridge, England: Cambridge University Press.

Chapman, G. B. (1991). Trial order affects cue interaction in contingency judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(5), 837.

Chouinard, M. M., & Clark, E. V. (2003). Adult reformulations of child errors as negative evidence. *Journal of Child Language*, *30*(3), 637–669.

Chuang, Y.-Y., Vollmer, M. L., Shafaei-Bajestan, E., Gahl, S., Hendrix, P., & Baayen, R. H. (2021). The processing of pseudoword form and meaning in production and comprehension: A computational modeling approach using linear discriminative learning. *Behavior Research Methods*, *53*, 945–976.

Culbertson, J., Gagliardi, A., & Smith, K. (2017). Competition between phonological and semantic cues in noun class learning. *Journal of Memory and Language*, *92*, 343–358. https://doi.org/10.1016/j.jml.2016.08.001

Dąbrowska, E. (2015). What exactly is universal grammar, and has anyone seen it? *Frontiers in Psychology*, *6*, 852.

Denistia, K., & Baayen, R. H. (2023). Affix substitution in Indonesian: A computational modeling approach. *Linguistics*, *61*(1), 1–32.

Denniston, J. C., Savastano, H. I., & Miller, R. R. (2000). The extended comparator hypothesis: Learning by contiguity, responding by relative strength. In R. R. Mowrer& S. B. Klein (Eds.), *Handbook of contemporary learning theories* (pp. 65–117). Mahwah, NJ: Lawrence Erlbaum Associates Publishers

Divjak, D., Milin, P., Ez-zizi, A., Józefowski, J., & Adam, C. (2021). What is learned from exposure: An error-driven approach to productivity in language. *Language, Cognition and Neuroscience*, *36*(1), 60–83.

Ellis, N. C. (2006). Selective attention and transfer phenomena in L2 acquisition: Contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics*, *27*(2), 164–194.

Ellis, N. C., Hafeez, K., Martin, K. I., Chen, L., Boland, J., & Sagarra, N. (2014). An eye-tracking study of learned attention in second language acquisition. *Applied Psycholinguistics*, *35*(3), 547–579.

Ellis, N. C., & Sagarra, N. (2010a). The bounds of adult language acquisition: Blocking and learned attention. *Studies in Second Language Acquisition*, *32*(4), 553–580.

Ellis, N. C., & Sagarra, N. (2010b). Learned attention effects in L2 temporal reference: The first hour and the next eight semesters. *Language Learning*, *60*, 85–108.

Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition*, *48*(1), 71–99.

Goldberg, A. E. (2019). *Explain me this*. Princeton, NJ: Princeton University Press. https://doi.org/10.1515/9780691183954

Harmon, Z., Idemaru, K., & Kapatsinski, V. (2019). Learning mechanisms in cue reweighting. *Cognition*, *189*, 76–88.

Harmon, Z., & Kapatsinski, V. (2017). Putting old tools to novel uses: The role of form accessibility in semantic extension. *Cognitive Psychology*, *98*, 22–44.

Hayes, B., & Londe, Z. C. (2006). Stochastic phonological knowledge: The case of Hungarian vowel harmony. *Phonology*, *23*(1), 59–104.

Hayes, B., Zuraw, K., Siptár, P., & Londe, Z. (2009). Natural and unnatural constraints in Hungarian vowel harmony. *Language*, *85*(4), 822–863.

Heckler, A. F., Kaminski, J. A., & Sloutsky, V. M. (2006). Differential cue salience, blocking and learned inattention. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *28*, 1476–1481.

Heitmeier, M., Chuang, Y.-Y., & Baayen, R. H. (2021). Modeling morphology with linear discriminative learning: Considerations and design choices. *Frontiers in Psychology*, *12*, 720713.

Heitmeier, M., Chuang, Y.-Y., & Baayen, R. H. (2023). How trial-to-trial learning shapes mappings in the mental lexicon: Modelling lexical decision with linear discriminative learning. *Cognitive Psychology*, *146*, 101598.

Hollis, G. (2019). Learning about things that never happened: A critique and refinement of the Rescorla-Wagner update rule when many outcomes are possible. *Memory & Cognition*, *47*, 1415–1430.

Hoppe, D. B., Hendriks, P., Ramscar, M., & van Rij, J. (2022). An exploration of error-driven learning in simple two-layer networks from a discriminative learning perspective. *Behavior Research Methods*, *54*, 2221–2251.

Hoppe, D. B., van Rij, J., Hendriks, P., & Ramscar, M. (2020). Order Matters! influences of linear order on linguistic category learning. *Cognitive Science*, *44*(11), e12910.

James, W. (1890). *The principles of psychology*, vol. 1. New York: Henry Holt.
Kamin, L. J. (1967). *Predictability, surprise, attention, and conditioning* (Technical Report No. 13). McMaster University Hamilton, Ontario, Canada. This paper was also presented at the Symposium on Punishment, May 1967, Princeton, NJ.
Kamin, L. J. (1968). Attention-like processes in classical conditioning. In *Miami symposium on the prediction of behavior: Aversive stimulation* (pp. 9–31). Oxford, OH: University of Miami Press.
Kapatsinski, V. (2010). Velar palatalization in Russian and artificial grammar: Constraints on models of morphophonology. *Laboratory Phonology*, *1*(2), 361–393.
Kapatsinski, V. (2018a). *Changing minds changing tools: From learning theory to language acquisition to language change*. Cambridge, MA: MIT Press.
Kapatsinski, V. (2018b). Learning morphological constructions. *The construction of words: Advances in construction morphology* (pp. 547–581). Cham, Switzerland: Springer.
Kapatsinski, V. (2023a). Defragmenting learning. *Cognitive Science*, *47*(6), e13301.
Kapatsinski, V. (2023b). Learning fast while avoiding spurious excitement and overcoming cue competition requires setting unachievable goals: Reasons for using the logistic activation function in learning to predict categorical outcomes. *Language, Cognition and Neuroscience*, *38*(4), 575–596.
Kordić, S. (1997). *Serbo-Croatian*. Number 148 in Languages of the World/Materials. Munich: Lincom Europa.
Kruschke, J. K. (2006). Locally Bayesian learning with applications to retrospective revaluation and highlighting. *Psychological Review*, *113*(4), 677.
Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13
Ladefoged, P., & Johnson, K. (2014). *A course in phonetics*. New York: Cengage Learning.
Lehiste, I., & Ivić, P. (1986). *Word and sentence prosody in Serbo-Croatian*. Cambridge, MA: MIT Press.
Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review*, *82*(4), 276.
MacLeod, B. (2015). A critical evaluation of two approaches to defining perceptual salience. *Ampersand*, *2*, 83–92.
MacWhinney, B. (2004). A multiple process solution to the logical problem of language acquisition. *Journal of Child Language*, *31*(4), 883–914.
Marcus, G. F. (1993). Negative evidence in language acquisition. *Cognition*, *46*(1), 53–85.
Marcus, G. F., Pinker, S., Ullman, M., Hollander, M., Rosen, T. J., Xu, F., & Clahsen, H. (1992). Overregularization in language acquisition. *Monographs of the Society for Research in Child Development*, *57*(4), 1–178.
Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), B101–B111.
McKenzie, C. R., & Mikkelsen, L. A. (2007). A Bayesian view of covariation assessment. *Cognitive Psychology*, *54*(1), 33–61.
Miller, R. R., Barnet, R. C., & Grahame, N. J. (1995). Assessment of the Rescorla-Wagner model. *Psychological Bulletin*, *117*(3), 363.
Miller, R. R., & Matzel, L. D. (1988). The comparator hypothesis: A response rule for the expression of associations. In G. H. Bower (Ed.), *Psychology of learning and motivation*, volume 22 of Advances in Research and Theory (pp. 51–92). New York: Academic Press.
Miller, R. R., & Witnauer, J. E. (2016). Retrospective revaluation: The phenomenon and its theoretical implications. *Behavioural Processes*, *123*, 15–25.
Mirković, J., Seidenberg, M. S., & Joanisse, M. F. (2011). Rules versus statistics: Insights from a highly inflected language. *Cognitive Science*, *35*(4), 638–681.
Nieder, J., Chuang, Y.-Y., van de Vijver, R., & Baayen, H. (2023). A discriminative lexicon approach to word comprehension, production, and processing: Maltese plurals. *Language*, *99*(2), 242–274.
Nieder, J., Tomaschek, F., Cohrs, E., & van de Vijver, R. (2022). Modelling Maltese noun plural classes without morphemes. *Language, Cognition and Neuroscience*, *37*(3), 381–402.
Nikolaev, A., Chuang, Y.-Y., & Baayen, R. H. (2022). A generating model for Finnish nominal inflection using distributional semantics. *The Mental Lexicon*, *17*(3), 368–394.

Nixon, J. S. (2020). Of mice and men: Speech sound acquisition as discriminative learning from prediction error, not just statistical tracking. *Cognition*, *197*, 104081.

Nixon, J. S., Poelstra, S., & van Rij, J. (2022). Does error-driven learning occur in the absence of cues? examination of the effects of updating connection weights to absent cues. In J. Culbertson, A. Perfors, H. Rabagliati, & V. Ramenzoni (Eds.), Proceedings of the 44th Annual Conference of the Cognitive Science Society (pp. 2590–2597). Austin, TX: Cognitive Science Society.

Nixon, J. S., & Tomaschek, F. (2021). Prediction and error in early infant speech learning: A speech acquisition model. *Cognition*, *212*, 104697.

Nixon, J. S., & Tomaschek, F. (2023). Introduction to the special issue emergence of speech and language from prediction error: error-driven language models. *Language, Cognition and Neuroscience*, *38*(4), 411–418.

Olejarczuk, P., & Kapatsinski, V. (2018). The metrical parse is guided by gradient phonotactics. *Phonology*, *35*(3), 367–405.

Olejarczuk, P., Kapatsinski, V., & Baayen, R. H. (2018). Distributional learning is error-driven: The role of surprise in the acquisition of phonetic categories. *Linguistics Vanguard*, *4*(s2), 20170020.

Pearl, L. (2022). Poverty of the stimulus without tears. *Language Learning and Development*, *18*(4), 415–454.

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). Psychopy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203.

Pinker, S. (1984). *Language Learnability and Language Development*. Cambridge, MA: Harvard University Press.

Port, R. F., & Leary, A. P. (2005). Against formal phonology. *Language*, *81*(4), 927–964.

Prasada, S., & Pinker, S. (1993). Generalization of regular and irregular morphological patterns. *Language and Cognitive Processes*, *8*(1), 1–56.

Pullum, G. K., & Scholz, B. C. (2002). Empirical assessment of stimulus poverty arguments. *The Linguistic Review*, *19*(1-2), 9–50.

R Core Team (2021),. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/

Ramscar, M. (2022). Psycholinguistics and aging. In *Oxford research encyclopedia of linguistics* (pp. 1–20). Oxford, England: Oxford University Press. https://doi.org/10.1093/acrefore/9780199384655.013.374

Ramscar, M., Dye, M., & Klein, J. (2013). Children value informativity over logic in word learning. *Psychological Science*, *24*(6), 1017–1023.

Ramscar, M., Dye, M., & McCauley, S. M. (2013). Error and expectation in language learning: The curious absence of mouses in adult speech. *Language*, 760—793.

Ramscar, M., & Port, F. R. (2016). How spoken languages work in the absence of an inventory of discrete units. *Language Sciences*, *53*, 58–74.

Ramscar, M., & Yarlett, D. (2007). Linguistic self-correction in the absence of feedback: A new approach to the logical problem of language acquisition. *Cognitive Science*, *31*(6), 927–960.

Ramscar, M., Yarlett, D., Dye, M., Denny, K., & Thorpe, K. (2010). The effects of feature-label-order and their implications for symbolic learning. *Cognitive Science*, *34*(6), 909–957.

Regier, T., & Gahl, S. (2004). Learning the unlearnable: The role of missing evidence. *Cognition*, *93*(2), 147–155.

Rescorla, R. A. (1988). Pavlovian conditioning: It's not what you think it is. *American Psychologist*, *43*(3), 151.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. A. H. Black, & W. F. Prokasy (Eds.), *Classical conditioning II: current research and theory* (pp. 64–99). New York: Appleton- Century-Crofts.

Rohde, D. L., & Plaut, D. C. (1999). Language acquisition in the absence of explicit negative evidence: How important is starting small? *Cognition*, *72*(1), 67–109.

Romain, L., Ez-zizi, A., Milin, P., & Divjak, D. (2022). What makes the past perfect and the future progressive? experiential coordinates for a learnable, context-based model of tense and aspect. *Cognitive Linguistics*, *33*(2), 251–289.

RStudio Team (2021). *RStudio: Integrated development environment for R*. Boston, MA: RStudio, PBC. http://www.rstudio.com/

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533–536.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928.

Sagarra, N., & Ellis, N. C. (2013). From seeing adverbs to seeing verbal morphology: Language experience and adult acquisition of L2 tense. *Studies in Second Language Acquisition*, *35*(2), 261–290.

Saiegh-Haddad, E., Hadieh, A., & Ravid, D. (2012). Acquiring noun plurals in Palestinian Arabic: Morphology, familiarity, and pattern frequency. *Language Learning*, *62*(4), 1079–1109.

Samuel, S., Cole, G., & Eacott, M. J. (2019). Grammatical gender and linguistic relativity: A systematic review. *Psychonomic Bulletin & Review*, *26*(6), 1767–1786.

Saxton, M. (2000). Negative evidence and negative feedback: Immediate effects on the grammaticality of child speech. *First Language*, *20*(60), 221–252.

Scholz, B. C., & Pullum, G. K. (2002). Searching for arguments to support linguistic nativism. *The Linguistic Review*, *19*(1-2), 185–223.

Shafaei-Bajestan, E., Moradipour-Tari, M., Uhrig, P., & Baayen, R. H. (2023). LDL-AURIS: a computational model, grounded in error-driven learning, for the comprehension of single spoken words. *Language, Cognition and Neuroscience*, *38*(4), 509–536.

Shafaei-Bajestan, E., Uhrig, P., & Baayen, R. H. (2022). Making sense of spoken plurals. *The Mental Lexicon*, *17*(3), 337–367. https://doi.org/10.1075/ml.22011.sha

Shahmohammadi, H., Heitmeier, M., Shafaei-Bajestan, E., Lensch, H. P., & Baayen, R. H. (2024). How direct is the link between words and images? *The Mental Lexicon*. 1–40. John Benjamins. https://www.jbe-platform.com/content/journals/10.1075/ml.22010.sha

Smolek, A., & Kapatsinski, V. (2023). Syntagmatic paradigms: learning correspondence from contiguity. *Morphology*, *33*, 287–334.

Song, H., & White, J. (2022). Interaction of phonological biases and frequency in learning a probabilistic language pattern. *Cognition*, *226*, 105170. https://doi.org/10.1016/j.cognition.2022.105170

Stout, S. C., & Miller, R. R. (2007). Sometimes-competing retrieval (socr): a formalization of the comparator hypothesis. *Psychological Review*, *114*(3), 759.

Subotić, L., Sredojević, D., & Bjelaković, I. (2012). *Fonetika i fonologija: ortoepska i ortografska norma standardnog srpskog jezika*. Novi Sad, Serbia: University of Novi Sad.

Szagun, G. (2011). Regular/irregular is not the whole story: The role of frequency and generalization in the acquisition of German past participle inflection. *Journal of Child Language*, *38*(4), 731–762.

Tomaschek, F., & Ramscar, M. (2022). Understanding the phonetic characteristics of speech under uncertainty-implications of the representation of linguistic knowledge in learning and processing. *Frontiers in Psychology*, *13*, 754395.

Tomasello, M. (2005). Beyond formalities: The case of language acquisition. *The Linguistic Review*, *22*, 183–197.

van de Vijver, R., & Baer-Henney, D. (2014). Developing biases. *Frontiers in Psychology*, *5*, 634.

van de Vijver, R., & Uwambayinema, E. (2022). A word-based account of comprehension and production of Kinyarwanda nouns in the discriminative lexicon. *Linguistics Vanguard*, *8*(1), 197–207.

van de Vijver, R., Uwambayinema, E., & Chuang, Y.-Y. (2024). Comprehension and production of Kinyarwanda verbs in the discriminative lexicon. *Linguistics*, *62*(1), 79–119.

Van Hamme, L. J., & Wasserman, E. A. (1994). Cue competition in causality judgments: The role of nonpresentation of compound stimulus elements. *Learning and Motivation*, *25*(2), 127–151.

van Rij, J., & Hoppe, D. (2021). *edl: Toolbox for error-driven learning simulations with two-layer networks*. R package version 1.0. https://CRAN.R-project.org/package=edl

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. New York: Springer-Verlag. https://ggplot2.tidyverse.org

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., & Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Widrow, B., & Hoff, M. E. (1960). Adaptive switching circuits. In *IRE WESCON convention record* (Vol. *4*, No. 1, pp. 96–104). California, USA: Western Electronic Show and Convention.

Wiese, R. (1996). *The phonology of German*. Oxford, England: Oxford University Press.

Yang, C. (2016). *The price of linguistic productivity: How children learn to break the rules of language*. Cambridge, MA: MIT Press.

Zsiga, E. C.. (2013) *The Sounds of Language: An introduction to phonetics and phonology*. Linguistics in the World. Oxford, England: Wiley-Blackwell.

Zsiga, E. C. (2020). *The phonetics/phonology interface*. Edinburgh, Scotland: Edinburgh University Press.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1: Proportions of correct choice for forced-choice contrasts in [−unlearning] (left) and [+unlearning] (right) in DM_condition.

Figure S2: Proportions of correct choice for forced-choice contrasts in [−unlearning] (left) and [+unlearning] (right) in PL_condition.

Figure S3: Learned connection weights for bigram cues to each morphological outcome for [−unlearning] (top) and [+unlearning] (bottom) in DM_condition.

Figure S4: Learned connection weights for bigram cues to each morphological outcome for [−unlearning] (top) and [+unlearning] (bottom) in PL_condition.

Table S1: Connection weights for the six bigrams with the highest activation for each morphological outcome ([−unlearning] group in DM_condition).

Table S2: Connection weights for the six bigrams with the highest activation for each morphological outcome ([+unlearning] group in DM_condition).

Table S3: Activation weights for the six bigrams with the highest activation within each morphological outcome ([−unlearning] group in PL_condition).

Table S4: Activation weights for the six bigrams with the highest activation within each morphological outcome ([+unlearning] group in PL_condition).

Table S5: Singular items paired with the single category blocks (learning I in [−unlearning]/learning II in [+unlearning]).

Table S6: Plural and diminutive items used in the respective single category phases (learning I in [−unlearning]/learning II in [+unlearning])

Table S7: Singular items used in the compounded morphological function blocks (learning II in [−unlearning]/learning I in [+unlearning])

Table S8: Diminutive-Plural items used in the compounded morphological function blocks (learning II in [−unlearning]/learning I in [+unlearning]).

Table S9: Novel plural and diminutive items used in the test phases.

Table S10: Diminutive and plural items repeated in the test phases.

Appendix A: Estimates for the pairwise interactions

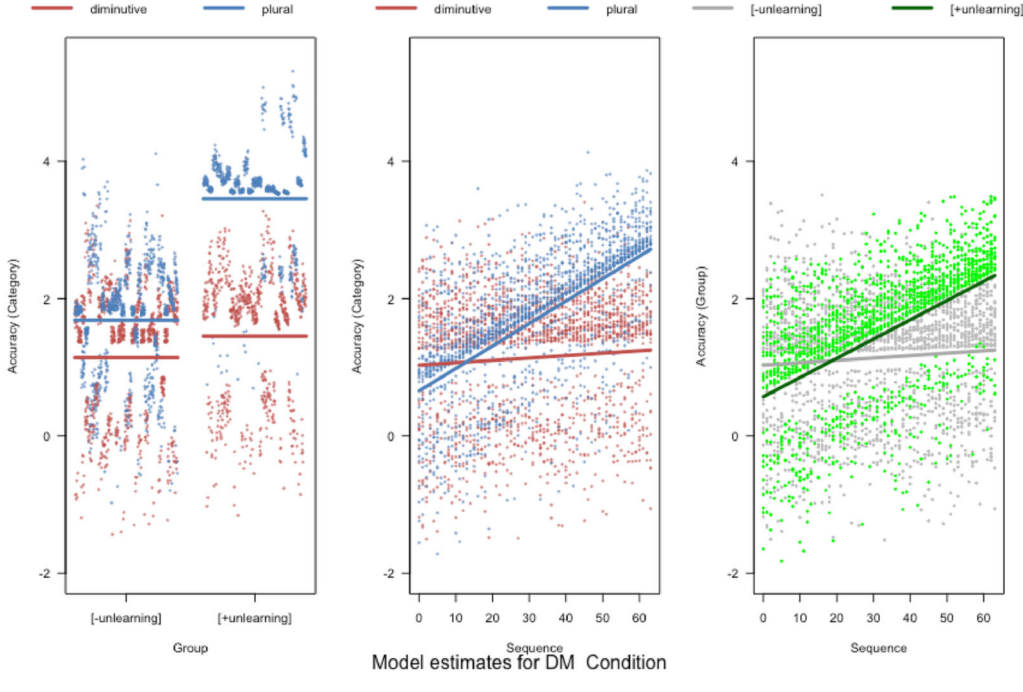

Fig. A.1. Model estimates for the pairwise interactions between `Group x Category`, `Category x Sequence`, and `Group x Sequence` in the DM_condition.

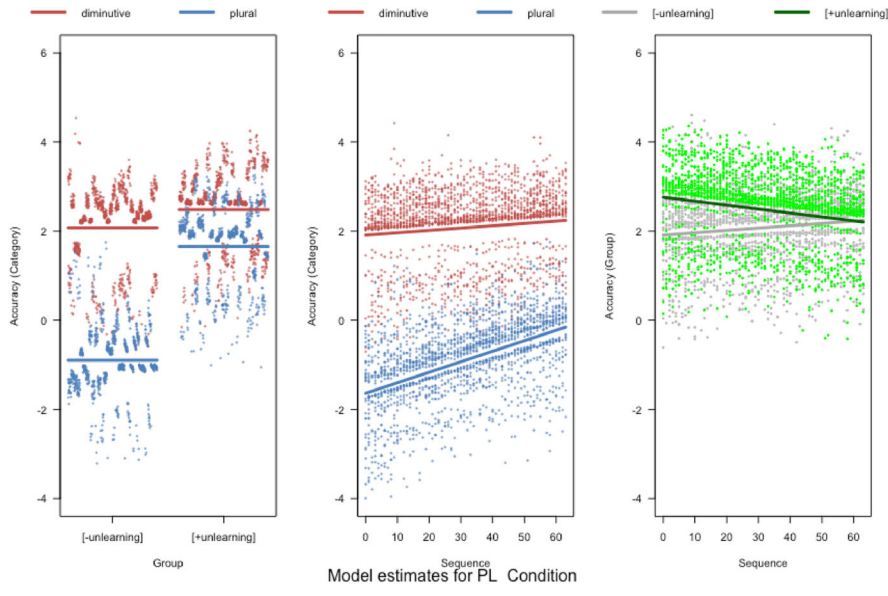

Fig. A.2. Model estimates for the two-way interactions between `Group x Category`, `Category x Sequence`, and `Group x Sequence` in the PL_condition.

Appendix B: Estimates for the three-way interactions

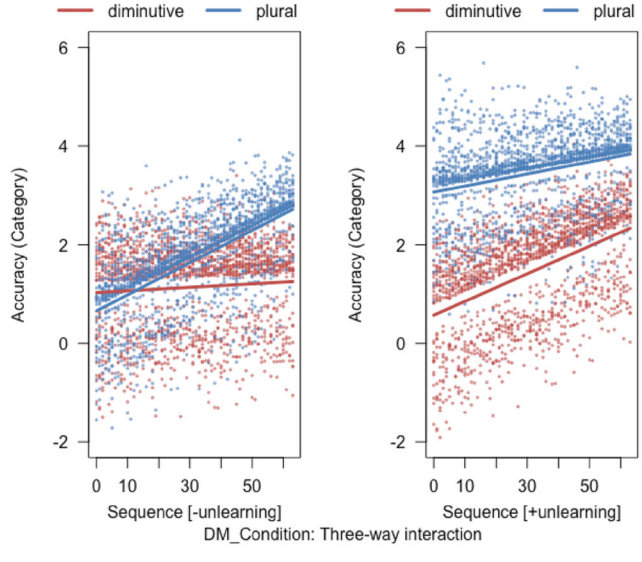

Fig. B.1. Model estimates for the three-way interaction between `Group x Category x Sequence` in the DM_condition.

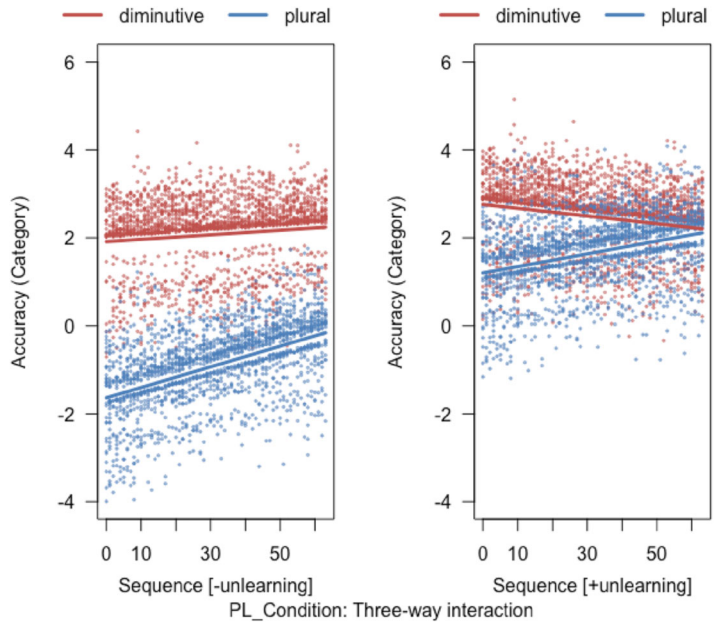

Fig. B.2. Model estimates for the three-way interaction between `Group x Category x Sequence` in the PL_condition.