COGNITIVE SCIENCE
A Multidisciplinary Journal

Check for updates

Cognitive Science 48 (2024) e70006

ISSN: 1551-6709 online
DOI: 10.1111/cogs.70006

Of Mouses and Mans: A Test of Errorless Versus Error-Based Learning in Children

Megan Waller,[a] Daniel Yurovsky,[a] Nazbanou Nozari[b,c]

[a]*Department of Psychology, Carnegie Mellon University*
[b]*Department of Psychological and Brain Sciences, Indiana University*
[c]*Cognitive Science Program, Indiana University*

Received 24 February 2024; received in revised form 1 October 2024; accepted 8 October 2024

Abstract

For both adults and children, learning from one's mistakes (error-based learning) has been shown to be advantageous over avoiding errors altogether (errorless learning) in pedagogical settings. However, it remains unclear whether this advantage carries over to nonpedagogical settings in children, who mostly learn language in such settings. Using irregular plurals (e.g., "mice") as a test case, we conducted a corpus analysis ($N = 227$) and two preregistered experiments ($N = 56$, $N = 99$), to investigate the potency of error-based learning as a mechanism for language acquisition in 3- and 4-year-old children. The results of the corpus analysis showed that incidental feedback after errors, in the form of caregivers' reformulations of children's errors, was relatively infrequent, had modest informational value, and was rarely used by children to correct their errors immediately. The following two experiments contrasted error-based learning with errorless learning, where the correct utterance was modeled for the child before a potential error was committed. The results showed that error-based learning was not always effective, and when it was, it was certainly not superior to errorless learning. Collectively, these findings question the extension of the benefits of error-based learning from pedagogical to nonpedagogical settings and define constraints under which one mechanism may be more beneficial to learning than the other.

Keywords: Language acquisition; Overregularization; Errorless learning; Error-based learning; Reformulation

Correspondence should be sent to Megan Waller, Department of Psychology, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA. E-mail: meganwal@andrew.cmu.edu

1. Introduction

How do children learn their native language? Research in cognitive science has often conceptualized this question as a problem of statistical inference over parent input (e.g., Pinker, 1979). The two critical elements in this view are the quality of the input that children receive (e.g., Cartmill et al., 2013) and their ability to make inferences based on that input (e.g., Gleitman et al., 2005; Nappa et al., 2009; Xu & Tenenbaum, 2007). Given the key role of these two factors, much research has been dedicated to exploring what information is available to children in input utterances (Gleitman, 1990), what constraints such input impose on their inferences (Markman & Wachtel, 1988), and what machinery do they use to perform these inferences (Smith & Yu, 2008). Modern theoretical frameworks have recently begun to incorporate children's own productions as part of the language input for learning. One such theory is the Usage-Based Theory of language learning, which focuses on how language is learned in the context of interactions (Tomasello, 2000). Under this framework, different inputs lead to individual differences in language use, and these differences in production will in turn lead to differences in external input (Behrens, 2021). Similarly, dynamic systems theory characterizes learning as a complex system where both the individual adapts their language to the environment, and the environment adapts to the individual (Van Geert & Verspoor, 2015). Both theories call for more research into this cyclic relationship: understanding how children's productions are shaped by input, how parents respond to their utterances, and then how children learn from these responses. However, most cognitive research on production has focused on learning phonological (Baese-Berk & Samuel, 2016, Keren-Portnoy et al., 2010) and articulatory-phonetic aspects of language (Goldstein & Schwade, 2008; Humphreys et al., 2010). Here, we extend this research to lexico-syntactic learning.

Children make many lexico-syntactic errors during language development (e.g., Hanley et al., 2016; Nozari & Omaki, 2022), and some errors are natural consequences of learning the structure of language. For example, as English-speaking children learn to extract morphological rules of their language for pluralization, it is reasonable to infer that "mouses" is the legitimate plural form of "mouse" and to produce that error. Young children frequently produce these overregularization errors up until the age of 6, where the rate of overregularization begins to steadily decrease (Marcus et al., 1992). Do these errors delay or facilitate learning? Errors can be a double-edged sword. If the child learns *from* the error with the help of feedback, it can boost their learning of the correct form. If, on the other hand, they learn the *error itself*, it can slow learning. These two perspectives formally map on to error-based and errorless learning and have been studied extensively in pedagogical contexts. This study aims to extend this research to learning in nonpedagogical settings.

1.1. Errorless versus error-based learning

Errorless learning emphasizes the associative mechanisms behind learning. Also known as Hebbian learning (Hebb, 1949), associative learning mechanisms propose that simultaneous activation of representations link them together. For example, seeing a picture of mice while hearing the word "mice" links the two. This manifests as learning, because the next time that the picture of mice is viewed, the word “mice” is activated through the connection previously formed between the two. According to this account, errors are detrimental to learning because they can lead to the formation of incorrect associations. For example, if viewing the picture of mice co-occurs with hearing the word “mouses,” associative learning can strengthen the incorrect pairing, leading to the erroneous retrieval and production of “mouses” next time a picture of mice is seen. Assuming children’s own production as input to their learning, their own errors would reinforce the incorrect associations. Therefore, to learn the correct form, the most effective method would be to avoid making errors and instead practice a modeled correct form.

In contrast to associative learning mechanisms, which emphasize learning from exposure and co-occurrence, error-based learning mechanisms emphasize the active participation of the learner in the learning process. An important element here is the idea of active information retrieval from the input. When seeing a picture of mice, learners use the input to activate an output, in the case of language learning, a word. As long as the learner retrieves the correct output, the predictions of error-based and errorless learning accounts are not at odds. But what if the learner retrieves the incorrect output, for example, “mouses”? Here, error-based learning and errorless learning theories make the opposite predictions. According to error-based learning, errors are good for learning, because they trigger corrective changes in the system, through the implementation of a simple learning rule, called the delta rule (Rumelhart et al., 1985). In a nutshell, the system takes an input and retrieves an output based on its current state. This “actual output” is then compared to a “target output” (sometimes called the teacher pattern), that is, what the system should have retrieved. Any discrepancy between the actual and target outputs serves as an error signal. This error signal is then used to adjust the connection weights between the input and output representations to increase the chance of retrieving the correct output in the future, while making the incorrect output less likely to be retrieved. Overall, since errors act as a learning signal, they are considered helpful for learning. Fig. 1 shows a schematic of an error-based model applied to the mice/mouses problem discussed above. As can be seen, the erroneous retrieval of “mouses” actually promotes the learning of the correct form “mice.”

It is worth pointing out that in perceptual tasks, the perceiver’s actual output is often a prediction (e.g., Rescorla & Wagner, 1972), whereas in production tasks, the actual output is a production, that is, the outcome of selecting one of several activated representations (see Nozari, 2018, for the problem of selection, and Oppenheim & Nozari, 2024, for mechanisms of selection processes in production). The “target output” is often provided by feedback, which can be explicit (as in most pedagogical settings) or implicit (as in many natural settings).

To recap, errorless learning predicts poorer learning when learners are allowed to make errors, because they could learn the incorrect associations, whereas error-based learning predicts better learning when learners make errors, because such errors trigger learning. But which account best explains how errors impact children’s learning? There is some support on either side. On the one hand, there are studies suggesting that errors are reinforced through production. For example, Humphreys, Menzies, and Lake (2010) showed that if speakers

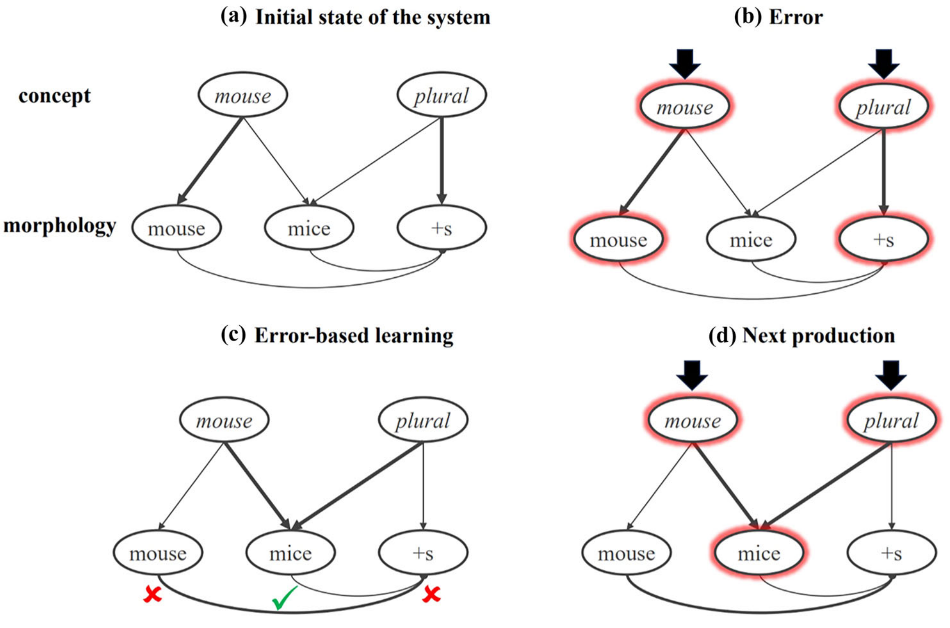

Fig. 1. A schematic of error-based learning in the pluralization task, using the example "mice."
Note. For simplicity, the network only shows the mapping of a concept (more than one mouse) to morphemes. Adding phonology and articulatory phonetics does not change the critical dynamics of error-based learning depicted here, since the locus of learning is mapping semantic to morphological representations. Arrows indicate excitatory connections and curves, inhibitory connections. The thickness of the arrows indicates the strength of the connections. (a) The initial state. The initial state of the system is informed by input throughout the child's life. It is likely that the child has had encounters with irregular plurals, so "mice" may already exist in the system, but since the majority of nouns encountered by the child are regular nouns, the initial state is likely to be biased toward regularization (stronger connections to mouse + s). (b) The first attempt. Given the bias toward regularization in the system, the first attempt is likely to produce "mouses." (c) Error-based learning. Feedback, such as a reformulation of the original utterance now using the correct word, identifies the correct response as "mice" (and not "mouse" + "s"). The system changes its connections in a way to increase the probability of retrieving "mice" next time the concept is activated. This readjustment obeys the delta rule: connections are altered proportional to the error signal and the activation of the supporting nodes. For correct responses, this alteration leads to strengthening, whereas for incorrect responses, it leads to weakening. (d) Next production. The changes to the connections resulting from error-based learning increase the likelihood of selecting "mice" over "mouses" next time the concept of several mice is activated. During errorless learning, steps (b) and (c) do not happen. The child simply produces "mice" (after a modeled utterance). This production could strengthen the weights between "mice" and its corresponding semantic features as shown in (d), although the pattern and magnitude of weight changes are not necessarily the same across error-based and errorless learning.

made a phonological error (e.g., "beg pet" instead of "peg bet"), they were four times more likely to repeat that error later in the experiment, suggesting that the error has been learned.

In contrast, there is some indirect evidence for the role of error-based learning in syntactic production in children (e.g., Fazekas et al., 2020; Lin & Fisher, 2017; Peter et al., 2015). The main idea behind these studies is exploiting a potential mismatch between an existing verb bias in the child's language system and the input they receive to investigate changes to production. Specifically, the studies use two structures, the prepositional-object structure (PO, e.g., "He showed/passed the book to her") and the double-object dative structure (DO, e.g., "He showed/passed her the book"), with two verbs with opposing biases. These studies demonstrate that when the verb is presented or trained in a context that violates internal predictions, that is, one that generates an error signal, children are more likely to shift to the structure they were exposed to (Fazekas et al., 2020; Lin & Fisher, 2017; Peter et al., 2015).

However, the most direct way to test the efficacy of errorless versus error-based account in language learning comes from learning paradigms that either allow participants to make errors and receive feedback (error-based mode) or model the utterance upfront to prevent speakers from making errors as much as possible (errorless mode). For example, Middleton et al. (2015) retrained individuals with aphasia in a picture naming task. In the error-based learning mode, participants were first encouraged to retrieve the picture's name. In the aphasic population, this often leads to errors. They then heard the correct response. In the errorless learning mode, they first heard the correct response and repeated it, thus making few errors. The authors found better learning in the error-based mode and replicated this effect in a similar study (Schuchard & Middleton, 2018; but see Fillingham et al., 2006, for different results. See also Fillingham et al., 2003, and Middleton & Schwartz, 2012, for reviews). While this study tests errorless and error-based learning in adults with aphasia, the simple and straightforward nature of the paradigm is well-suited for testing these two learning theories in children. In a similar design, Saxton (1997) taught children the present and past tense of novel irregular verbs (e.g., present tense: "pell"—past tense: "pold"). In the errorless condition, children first heard the correct form and then repeated it. In the error-based condition, they first attempted to produce the form and then heard the correct form. In keeping with Middleton et al. (2015), Saxton (1997) found superior learning in the error-based condition. Saxton et al. (1998) replicated this result and extended it to a 5-week learning period.

These studies make a convincing case for the superiority of error-based learning in pedagogical settings and its extension to young children's learning of morphosyntactic forms. It is tempting to extrapolate these findings to claim that the same error-based learning mechanism is the most potent mechanism for driving language acquisition in nonpedagogical settings. However, an important feature of pedagogical settings is that participants' primary goal and focus is learning words. As such, feedback, which is provided regularly and explicitly in such settings, can be assumed to be fully processed to trigger learning. In contrast, in most nonpedagogical settings, where most of children's language learning actually takes place, learning is incidental, that is, language exposure occurs in the context of an activity with a goal different that teaching the child a new word. It thus remains unclear whether feedback is reliable and fully processed in such settings. If so, one could expect a clear extension of the advantage of error-based learning mechanisms from pedagogical to nonpedagogical settings. Conversely, unreliable feedback or feedback not used properly would cast doubt on the superiority of error-based learning mechanisms in driving learning in nonpedagogical settings.

1.2. Feedback processing during language learning in children

When learners make an error, corrective feedback should help with repairing that error and learning the correct target. Critically, feedback is most useful when paired closely in time with the error (Goldstein & Schwade, 2008; Roseberry et al., 2014; see Masek et al., 2021, for review). Aside from timing, two factors call for studies designed specifically to understand feedback processing as children learn language in naturalistic environments: (a) the form and frequency of feedback that children receive and (b) children's ability or willingness to process that feedback at the service of learning.

Unlike pedagogical settings, where errors are almost always followed by explicit feedback which unequivocally point out a need for correction, natural language learning in children does not frequently contain explicit error signals. And when it does, such error signals may be disguised, making it harder for the child to register them. For example, when children produce overregularizations such as "mouses," parents rarely explicitly correct these errors, for example, by producing utterances like "No! You say mice." (Hirsh-Pasek et al., 1984). Rather, they sometimes provide implicit negative evidence by replaying the child's utterance, but changing the error to the correct form, for example, "Yes, those are mice!" (Chouinard & Clark, 2003; Hirsh-Pasek et al., 1984). This is called a *reformulation*, and it provides a subtle and indirect signal to the child that they may have committed an error (Clark, 2020). It is unclear how prevalent reformulations are in naturalistic conversation. Corpus analyses have reported varying rates. For example, Chouinard and Clark (2003) and Farrar (1992) found that parents reformulated almost half of their children's errors, whereas other studies have found much lower rates (e.g., 15%, Bohannon & Stanowicz, 1988; 21%, Hirsh-Pasek et al., 1984).

Let us assume, for the moment, that reformulations are frequent. Do children actually use them for correcting their errors? There is some corpus evidence that reformulations can be a useful learning signal and that children attend to them. For instance, Chouinard and Clark (2003) demonstrated that parents are more likely to reformulate children's errors than to repeat correct utterances. They also reported that children frequently repeat, acknowledge, or explicitly reject reformulations when their parents misunderstood them. In addition, children have been shown to repeat their parents' utterances more after reformulations than noncorrective repetitions (Farrar, 1992). Finally, grammatical forms of irregular words are more common in children's speech after receiving reformulations than other noncorrective responses (Nelson et al., 1973; Saxton, 2000). But these findings are not uncontested. For example, in a corpus analysis conducted by Morgan et al. (1995), while the authors found similarly high rates of reformulations as Chouinard and Clark (2003), they reported that children were not more likely to correct themselves after reformulations compared to responses that did not contain a correction or replay, suggesting children may not detect or make use of these implicit corrections (see also Morgan & Travis, 1989). One caveat is that all these corpus studies have had a small sample of children, with as few as one child (Saxton, 2000), and at most 27 children (Nelson et al., 1973). In short, reformulations are precisely the kind of feedback that is expected to drive error-based learning in nonpedagogical settings. However, reports of the rate and efficacy of reformulations vary substantially, making it difficult to draw clear conclusions about the impact of children's errors on the process of language development.

1.3. The current study

This study examines whether error-based learning, which has consistently been shown to benefit learning more in pedagogical settings, is more beneficial for language development than errorless learning in nonpedagogical settings. We use the overregularization of plural morphology as a test case because it is a common error in children under the age of 6, and neurotypically developing children recover from them seemingly spontaneously by the age of 10 (Marcus et al., 1992; McClelland et al., 1987; Ramscar et al., 2013). Since few parents provide explicit instructions on inflectional morphology to their young children (Hirsh-Pasek et al., 1984), these errors are an excellent test case for studying the role of incidental feedback in learning.

Take two children both of whom start with no knowledge of irregular pluralization (a state that every child experiences at some point in their life). Both children live in an environment where the correct plural forms are spoken and thus receive them as input (again, a fair assumption for the majority of children). Child 1 is a talkative child who produces many overregularization errors, and thus receives environmental input as feedback to their errors. Child 2 is a taciturn child, who rarely attempts to produce a plural whose form they are unsure of. Thus, the environmental input received by this child is not in the form of feedback to an error. Both children eventually learn the correct plural forms. The question is which one shows more efficient learning?

Ideally, one would study this in children's natural environments. This was the motivation behind our corpus study. The corpus analysis allowed us to assess, using a larger sample size than used in prior studies, whether incidental feedback after errors was prevalent and informative, and whether children used it to correct their errors. However, while the corpus study allows us to look at learning in children similar to Child 1 in our example, it is impossible to get a baseline for children similar to Child 2 (i.e., with Child 1, we have a baseline of when they did not know the correct form and can evaluate whether they did or did not learn from feedback, but with Child 2, we do not know whether they did or did not know the correct form to begin with). Thus, the corpus study provides useful but incomplete information. Even if we somehow had access to the internal states of Child 2 in the corpus, there may have been a confound: children who make errors may elicit feedback and thus get more exposure to the correct forms. If so, Child 1's advantage could be due to greater exposure to the correct form and not due to the impact of errors on learning. Finally, Child 1 and Child 2 may have different temperaments that affect learning for reasons other than cognitive principles of learnings. To address these issues, we conducted two empirical studies which (a) equated input for both cases, and (b) reduced the influence of personality by randomly assigning children to learning conditions (Experiment 1) and by using a within-subject design (Experiment 2).

The experimental design followed the classic errorless versus error-based paradigm design employed by Saxton (1997) and Middleton et al. (2015), embedded in a "dream-machine" game. In Experiment 1, using a between-subject design, the child provided counts + nouns (e.g., "one snake and two snowmen/snowmans") to help the experimenter find their dream. In Experiment 2, in a within-subject design, the task was simplified by dropping the counting ("snakes and snowmen/snowmans"). In the errorless (Modeled) condition, the experimenter first produced an utterance containing the critical nouns and the child then repeated it. In the error-based (Reformulated) condition, the child first attempted to produce the utterance, and then heard the correct utterance from the experimenter. In a pre versus post design, we compared the change to children's pluralization errors in the two conditions.

If, similar to pedagogical settings, error-based learning is also a more effective mechanism of language learning than errorless learning in nonpedagogical settings, we expect to replicate the findings of Saxton (1997) and Saxton et al. (1998), namely, better learning in the error-based than the errorless condition. If, on the other hand, children cannot effectively employ reformulations to trigger error-based learning mechanisms, then learning in the errorless condition should be at least as good or better than the error-based condition.

1.4. Data availability

The OSF repository containing the data and analysis files for all studies can be found here: https://osf.io/bw8zh/?view_only=c923e0c88d954d61920009067dff8dc4.

2. Corpus analysis

Reformulation rates estimated in prior work vary widely, from 15% to 48% of errors (Bohannon & Stanowicz, 1988; Chouinard & Clark, 2003; Farrar, 1992; Hirsh-Pasek et al., 1984; Morgan et al., 1995). To get a more precise estimate for overregularization errors like "mice," we investigated all children available in the North American and British English corpora of the Child Language Data Exchange System (CHILDES; MacWhinney, 2000). We also measured children's immediate uptake of corrections provided in reformulations.

2.1. Method

We searched CHILDES for irregular plural items that children acquire before the age of 6, as the past literature indicates that the majority of overregularization errors (e.g., "goed" for "went") have disappeared by this age (Marcus et al., 1992). This was formally done by selecting items with an Age of Acquisition (AoA) of lower than 6 according to Kuperman et al. (2012). Table 1 shows these items, along with their AoAs and lexical frequencies (Brysbaert & New, 2009). From CHILDES, we sampled from both the English-United Kingdom and the English-North American corpora any instance in which a child said a plural form of our target irregular plurals, along with the two utterances that followed, in order to capture possible reformulations and the child's response to the reformulation. The correct form and two common forms of overregularization errors were registered: adding an "s" on either the singular form ("mouses") or the plural form ("mices"). While this may not capture every time the child attempts a plural form, this sample likely covers the most common cases. We then only selected interactions in which someone other than the target child responded to the initial utterance. This resulted in a sample of 3397 total correct and incorrect utterances across 227 children (11.82−133.46 months; M_{age} = 40.17 months; SD_{age} = 16.17 months). Considering most initial errors (94.2%) were from children up to and including the age of 4, and nearly all the reformulations in this sample (see Table 2), the following analyses

Table 1
Log10 word frequency (per one million words) and age of acquisition (AoA) of target words in the corpus analysis

Word	Singular frequency	Plural frequency	Mean AoA	SD AoA
man	**4.97**	4.28	3.11	1.24
woman	**4.34**	3.99	4.95	2.19
child	3.91	**3.95**	5.15	1.70
foot	3.52	**3.79**	3.44	1.38
tooth	2.75	**3.38**	3.61	1.09
mouse	**2.99**	2.52	4.94	1.76
goose	**2.82**	1.91	5.15	2.21
leaf	2.42	**3.36**	4.60	2.37
wolf	**3.01**	2.48	4.50	1.79
wife	**4.25**	2.90	5.67	3.20
knife	**3.37**	2.54	4.15	1.80
half	**4.00**	1.75	5.45	1.77
snowman	**1.99**	0.70	3.47	0.90
policeman	**2.77**	2.30	4.44	2.01
mailman	**2.17**	1.04	4.72	1.60
newsman	**1.32**	1.20	5.72	2.22
postman	**2.11**	0.77	5.22	2.10
fireman	**2.17**	2.04	5.05	1.22
fisherman	**2.34**	2.04	5.16	1.77

Note. Bold numbers make it easier to spot singular-dominant and plural-dominant nouns.

Table 2
Adult replays and reformulations of children's incorrect utterances by age

Age (years)	Total children	Total utterances	Total errors	Replay of errors	Reformulation of errors	Move on from errors
0	**1**	**1**	**0**	**0**	**0**	**0**
1	**31**	**254**	**6**	**1 (16.7%)**	**4 (66.7%)**	**1 (16.7%)**
2	**87**	**1265**	**115**	**23 (20%)**	**32 (27.8%)**	**60 (52.2%)**
3	**55**	**543**	**47**	**6 (12.8%)**	**7 (14.9%)**	**34 (72.3%**
4	**41**	**355**	**29**	**2 (6.9%)**	**7 (24.1%)**	**20 (68.9%)**
5	5	56	1	0	0	1 (100%)
6–10	14	95	0	0	0	0
Sum (ages 0–4)	**151**	**2418**	**197**	**32 (16.2%)**	**50 (25.4%)**	**115 (58.4%)**
Sum total	160	2569	198	32 (16.2%)	50 (25.3%)	116 (58.6%)

Note. While our original corpus search included children of all ages, most errors were made by children aged 4 years old and under. Therefore, children ages 5 and older are not included in the subsequent analyses. The data included in these analyses are highlighted in bold. Additionally, the subtotal and total children listed here do not directly add to the total from each row, because some children were included in the corpus over the course of multiple years. The numbers reported in the totals are the counts of unique children across the respective age groups.

focused on 4-year-olds and younger, which make up 94.1% of the total utterances (correct or incorrect) in our original sample.

Since our interest is in characterizing the feedback provided by caregivers in children's natural environments, we then only selected interactions in which an adult caregiver [tagged in the corpus as "mother," "father," "grandmother," or "grandfather"] responded to the initial utterance. Utterances from other children or those obtained in experimental settings were not included (577 utterances). This left 2418 utterances, of which 197 utterances included an overregularization error (8.2% error rate). Specifically, two types of caregiver response were registered: (a) a ***replay***, where the caregiver simply repeated the child's plural form (e.g., "mice" → "mice"/ "mouses" → "mouses"), and (b) a ***reformulation***, where the adult provided a form different from the plural in the child's utterance (e.g., "mouses" → "mice," or much less commonly "mice" → "mouses"). Adults' utterances did not have to exactly repeat the form of the child's utterance in order to be included; only the target item needed to be said for the response to be considered a reformulation or replay. Consistent with prior literature (Hirsh-Pasek et al., 1984), most reformulations in the corpus were implicit, except for two utterances in which the adult said, "you mean [correct noun]." We also registered whether children replied with the correct form after receiving a reformulation.

2.2. *Results*

Table 2 shows the adult replays and reformulations of children's errors broken down by age. Total number of utterances and errors made by children are shown together with three types of adult responses to children's errors specifically: replays (repeats of a child's error), reformulations (providing a correction to the child's error), and move-on responses (answers that did not contain the child's error or an altered form of it). We are specifically interested in replays and reformulations. Fig. 2 summarizes the average rates of these two responses across individual children in the corpus, this time showing rates of responses to correct utterances for comparison.

Overall, when children produced the correct irregular plural (e.g., "mice"), adults replayed the plural form about one-third of the time (855 out of 2221 correct utterances, 38%). However, when children produced an error like "mouses," adults also sometimes replayed their overregularizations (32 out of 197 incorrect utterances, 16%), although the latter occurred at significantly lower rates (t (2416) $= 6.26$, $d = 0.47$, $p < .001$, two-sample t-test). As for reformulations, when children produced an error, adults reformulated their utterances about a quarter of the time (50 utterances out of 197 incorrect utterances, 25%). Reformation of the correct response also occasionally occurred (3 out of 2221 correct utterances, $< 1\%$), but at significantly lower rates (t (2416) $= -26.3$, $d = 1.95$, $p < .001$, two-sample t-test). All three of these cases were instances of an adult adding "-s" to the correct irregular form. Twice an adult replied with the plural form "childrens" after the child said "children," and once an adult replied with the form "feets" after the child said "feet." Overregularization errors were less frequent for plural-dominant targets (child, foot, leaf, tooth; 38% of the 197 errorful utterances), which aligns with work by Biedermann et al. (2013) showing longer response latencies for naming singular-dominant nouns in adults. Importantly, however, parental

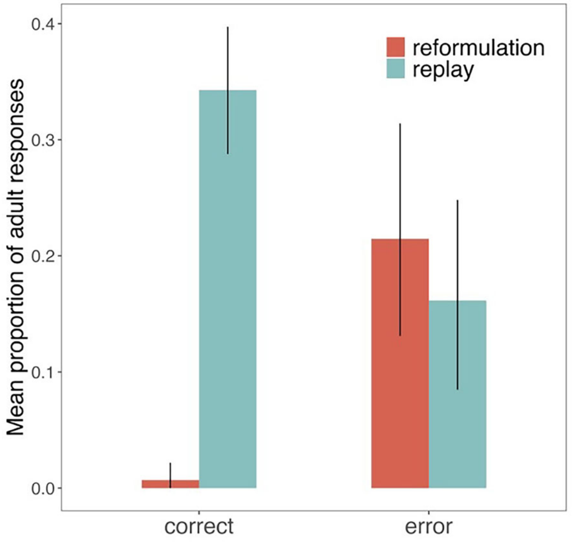

Fig. 2. Adults' average rates of replay and reformulation to individual children's correct responses and errors. *Note*. Values show mean of subject means with 95% confidence intervals estimated by nonparametric bootstrapping.

reformulation rate was comparable for overregularization errors made on plural-dominant (27%) and singular-dominant (25%) errors, indicating that parents do not respond differently to items with different frequencies in plural use.

These results show that reformulations are infrequent relative to the number of errors that children make, and they are less common than replays of correct plural utterances. However, the distributions of replays and reformulations across correct and error utterances are strikingly different: reformulations occur almost exclusively after error responses, whereas replays occur after both types of responses, albeit with a greater frequency after correct responses. The different distributions of these two response types, together with the different base rates of correct and error utterances in children, may cause reformulations and replays to carry different amounts of information for signaling errors. To assess this, we calculated information gain (IG), which is a measure of the reduction of entropy in a dataset when given new information. Entropy refers to the probability distribution of a variable, so in the case of accuracy, it captures the likelihood of a correct utterance. High entropy values indicate that correct and incorrect utterances are more or less equally likely, whereas low entropy points to the greater probability, or expectedness, of one utterance type. IG measures whether adding another variable, in this case, response type (reformulations or replays), can help reduce the entropy, or make a certain kind of utterance more predictable. Larger IG values indicate that a given response type provides better information for signaling errors. Information Gain was calculated as:

$$IG\,(A, R) = Entropy\,(A) - Entropy\,(A, R)\,,$$

where A is child's accuracy (correct, error), and R is adult's response type. Response type itself was divided into replays (yes, no) and reformulations (yes, no). Overall, this formula takes the dataset with only information of the child's accuracy and subtracts the dataset when you know both accuracy and response type, to understand how much information can be gained by the response type variable. IG was calculated separately for replays and reformulations. Generally, entropy of a variable such as X with c levels is computed as:

$$Entropy\,(X) = \sum_{i=1}^{c} -p_i \log_2 p_i,$$

where p_i indicates the probability of each level of factor X. Given that accuracy has two levels (correct, error) in our data,

$$Entropy\,(A) = -\left(p_{Correct} * \log_2 p_{Correct} + p_{Error} * \log_2 p_{Error}\right).$$

Similarly, entropy of (X, Y) is generally computed as:

$$Entropy\,(X, Y) = \sum_{c \in Y} P\,(c)\,E\,(c)\,,$$

where $P(c)$ and $E(c)$ denote the probability and entropy of each level of variable Y, respectively. In our case, this is computed once for replays and once for reformulations. For example,

$$Entropy\,(A, Replay) = p\left(Replay_{Yes}\right) * Entropy\left(Replay_{Yes}\right) + p\left(Replay_{No}\right) * Entropy\left(Replay_{No}\right).$$

Using this method, $IG(A, Replay)$ and $IG(A, Reformulation)$ were 0.013 and 0.072 bits, respectively (for a more thorough walk through of the calculations, see Appendix A). To put the effect of information in more concrete terms, consider the entropy of the dataset without considering the informativeness of reformulations, that is, $Entropy(A)$, which was 0.407. This number reflects a split of children's response accuracy into 91.9% correct and 8.1% errors. Ideally, the $Entropy(A)$ would be driven down to 0, reflecting 100% expected responses. With the information provided by reformulations, the entropy of the dataset, that is, $Entropy(A, R)$, was decreased to 0.335, reflecting a split of 93.8% versus 6.2% probability of correct responses. This means that knowing whether the adult did or did not provide a reformulation increases the probability of accurately identifying whether an utterance was correct or not by about 2%. In short, the IG of neither replays nor reformulations is particularly high, but between the two, reformulations are a more informative signal for learning and could cause a modest reduction in error if used appropriately by children.

But do children actually use reformulations to correct their errors? To answer this question, we looked at children's responses to adults' reformulations. Of the 50 reformulations to incorrect utterances, only seven replies contained any attempts on children's part to utter the target word. Of those, five were repetitions of children's original error (despite having been corrected through reformulations), leaving only two instances where the child successfully corrected their error through the reformulation. In other words, reformulations only led to immediate error correction in 4% of the cases in the corpus.

2.3. Discussion

Three findings were noteworthy from the corpus analysis: (1) First, we found that reformulations are not very frequent. After making an overregularization error, our analysis returned a reformulation rate of 25% in children up to and including age 4, where these errors are most common. Compared to other studies, this number is closest to that reported by Hirsh-Pasek et al. (1984), where they found a rate of 21%. Other studies have found lower rates, such as 15% from Bohannon and Stanowicz (1988), while others have found higher rates ranging from 41% to 48% (Chouinard & Clark, 2003; Farrar, 1992; Morgan et al., 1995). Although there are fewer errors per child, the sample size of the current corpus analysis is much larger than that reported in any of the prior studies and may thus better capture the variability in the population. (2) Second, we found that despite their infrequency, reformulations have higher informational value than replays. This is because parents almost never reformulate a child's correct utterance. Therefore, when they do, it can potentially provide a useful signal that the utterance needs to be changed. But even though they are better signals than replays, the informational value is still modest. (3) Children do not seem to use reformulations very effectively, at least over short periods of time. When children reattempted producing the target word after a reformulation, they repeated their own errors more frequently than the adult's reformulation. This finding matches the finding of Morgan et al. (1995) who also reported that children did not correct themselves more after a reformulation. In short, while reformulations could be used as a learning signal, there was no strong indication in this corpus analysis that they were.

Thus far, the results of the corpus analysis do not provide strong support in favor of effective learning from reformulation, and more generally error-based learning mechanisms with incidental feedback. But this might be because reformulations are not very frequent in the corpus. Although this finding is important per se, it may undermine the efficacy of error-based mechanisms when feedback is consistently provided. We investigated this issue in two experiments.

3. Experiment 1

Children played an interactive game with the experimenter in which they helped the experimenter fix a broken dream machine, so she could better remember her dreams. On each trial, a number of animals appeared in dream clouds and children were instructed to describe the content of the cued cloud (see Fig. 3). For example, they would say "one snake and two snowmen/snowmans." The experiment had two conditions. In the ***Reformulated condition***, children first made an attempt at producing the sentence and then heard the correct version produced by the experimenter. This condition reflects error-based learning opportunities. In the ***Modeled condition***, the experimenter first partially modeled the utterance, for example, by saying "I remember a snake and some snowmen." The child then provided the complete utterance with the counts. This condition represents errorless learning. To equate input between the two conditions, in the Reformulated condition even if the child produced the

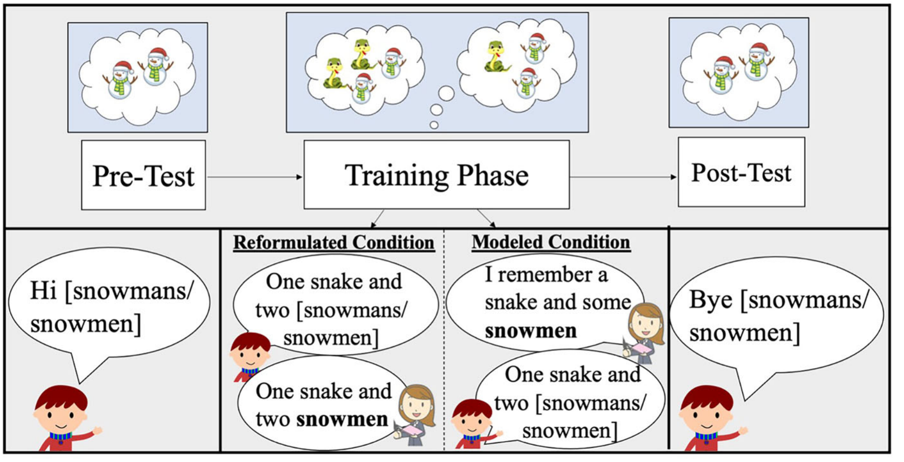

Fig. 3. Diagram of experiment flow for Experiment 1.

correct item, the experimenter would still provide the correct response. This way, all children heard the correct form the same number of times. The only difference was the timing; in the Reformulated condition, the correct utterance came *after* an error. In the Modeled condition, *before* a potential error.

3.1. Method

3.1.1. Participants

The results of the corpus analysis were used to guide participant selection. As shown in Table 2, the majority of overregularization errors were committed by children of 2−4 years of age, with this kind of error almost completely disappearing after age 4. Our pilot testing indicated that 2-year-old children were too young to participate in our experimental game (cf., Arnon & Clark, 2011, for other kinds of tasks). However, children as young as 3 years of age could understand and play the game with the instructor. For this reason, we focused on 3- and 4-year-old children.

We did not have a prior effect size corresponding to incidental learning from reformulations or to the comparison of learning rates between reformulating and modeling, to use as our target effect size for estimating the necessary sample size. We, therefore, prespecified a sample size of 60 children (Preregistration: https://osf.io/5edp7?view_only=c923e0c88d954d61920009067dff8dc4) comparable to other studies (Arnon & Clark, 2011; Fazekas et al., 2020; Lin & Fisher, 2017). To account for attritions, we recruited 66 3-and 4-year-olds from the Psychology Department database and through community advertisement. The only inclusion criteria were age, that children were neurotypically developing, and were English speakers. All children participated online over Zoom, and families received a $10 Amazon gift card for their participation. Six children did not complete the game, and one child was removed for experimenter error. The final sample consisted of 59 children (35 biologically female, 3.09−4.96 years, M_{age} = 4.00, SD_{age} = 0.50). The average age of children in the two experimental conditions was not significantly different from one another, as compared by an independent sample *t*-test (t (57) = 1.63, $p = .11$). The study was approved by the university's Institutional Review Board (IRB).

3.1.2. Materials

We also used the corpus analysis to select the experimental materials. Children were tested on eight irregular plurals (mice, geese, leaves, teeth, feet, men, women, snowmen). These items were selected because they were the most frequent in the corpus, with the exception of "snowman." "Snowman" was selected instead of "fireman" (which was more common), because its mean age of acquisition was lower, and it is more familiar to children from the local area, which is a snowy region. In addition, eight regular plurals (snakes, frogs, pigs, cats, dogs, spoons, forks, turtles) were included as fillers. These items were also high-frequency and acquired before 6 years (see Appendix B for frequencies). The images corresponding to these words were acquired from Google Images and the Public Domain Vectors website.

3.1.3. Procedure

Informed consent was obtained by parents before beginning the game. On Zoom, children were told they would play a game to help the experimenter fix their dream machine. The game lasted about 15 min on average. Each session had four phases: a *pre-test*, *two blocks of training trials*, and a *post-test* (Fig. 3). In the pre-test, children were instructed to greet each set of pictured entities, for example, by saying "Hi snowmans/snowmen!" to serve as an evaluation of the children's prior irregular word knowledge. If children did not know the name or used the wrong word (e.g., saying "girl" instead of "woman"), the experimenter provided the singular form of the noun and asked the child to try again. Half of the pictures were target irregular nouns (mice, geese, leaves, teeth, feet, men, women, snowmen). The other half were filler regular nouns. The regular plurals were used as fillers as to not direct the child's attention to the irregular nature of the target items.

After the pre-test came two blocks of training, which contained the study manipulation. Children were told that the dream machine could not tell which of two dreams was correct, and that they could help the experimenter by counting and naming the objects that distinguished the experimenter's dream (the dream cloud connected to the trail of bubbles) from another similar dream (Fig. 3). In each trial, the correct dream cloud blinked, and the child described the content of the cloud to the experimenter with a conjoined noun phrase, for example, "one snake and two snowmen/snowmans." If the child named the wrong dream or counted incorrectly, the experimenter reminded them to focus on the right dream and asked them to try again.

Each of the irregular plural nouns occurred once within each block. The first block had 16 trials in which every noun appeared in its singular and plural form. Trials were $\frac{1}{4}$ singular-singular, $\frac{1}{4}$ regular plural-irregular plural, and $\frac{1}{2}$ a singular-plural with a combination of regular and irregular nouns. Two irregular plurals never appeared together within the same trial. The second block had only 10 trials, removing some filler trials to reduce the duration of the game and thus maintain children's attention. There were two fixed orders of trials counterbalanced across children.

Before completing the experimental trials, the experimenter modeled the task for children and instructed them to repeat after her. Children then practiced the game with two trials that only contained regular plurals and received feedback on their performance until they were able to do the task correctly. They then completed the two training blocks with breaks in between. Children were randomly assigned to one of the two experimental conditions in the training phase. In the Reformulated condition, children first produced a conjoined noun phrase and then heard the correct form from the experimenter. If children did not make a mistake, the experimenter replayed the correct response (see Fig. 3). In the Modeled condition, the experimenter said each correct form before the child's turn (e.g., "I remember a snake and some snowmen"), thus intercepting most of the errors before they surfaced. The experimenter did not give any feedback after children's productions. After finishing a trial, the experimenter then pressed a button to start a new trial. Finally, during the post-test, the children said "goodbye" to each of the pictured entities again (e.g., "Bye-bye snowmen!"), to test for any change in their irregular noun production after the intervention.

3.1.4. Data preprocessing

Zoom sessions were recorded and then transcribed by three coders, who were native speakers of American English. The coders marked whether the child provided the right number (numerical error), whether the child used the correct root noun to refer to the object (lexical error), and whether the child used the correct plural form (pluralization error). If the child did not provide any number or word during a trial, it was coded as a numerical or lexical error, respectively. Of the 59 videos, 12 videos were coded independently by a pair of coders to detect the reliability of their coding. Coder judgments were reliably similar across all three error types (numerical: Cohen's $\kappa = 0.99$, lexical: $\kappa = 0.86$, and plural: $\kappa = 0.93$). The coders then met to reconcile any disagreements they had regarding the responses of those participants. Disagreements remained on 15 trials, and those were excluded from the analysis.

We intended to remove children who did not follow directions, and since that judgment could become subjective, we instantiated a rule of lexical error rate of over 20% (note this is independent of plural errors). Three children were removed because their error rate on producing the correct lexical items exceeded 20%. This resulted in a sample of 56 children (34 biologically female, 3.19−4.96 years, $M_{age} = 4.04$, $SD_{age} = 0.49$). The average age of children in the two experimental conditions was still not significantly different from one another, as compared by an independent sample *t*-test (t (54) = 1.14, $p = .26$). Of these participants, 30 were in the Modeled condition, and 26 were in the Reformulated condition. In addition, any trials in which the child used the wrong lexical item were also removed from analyses. If the child counted in a way that changed the noun from singular to plural or from plural to singular, those trials were also removed (e.g., counting one mouse instead of the two mice pictured). Finally, 15 trials on which the two coders did not reach a consensus were excluded. These measures combined removed 8.04% of trials.

Before presenting the results, a sample size issue needs to be addressed. Since we excluded three children based on their performance, the target sample size is four subjects lower than the preregistered 60. To assess the severity of this issue, we conducted a power analysis based on the only available effect size from a previous study on learning from reformulations, that is, Saxton (1997), using G*Power 3.1 (Faul et al., 2009). The effect size in that study was quite large (Cohen's *d* of 5.2), and given an alpha of 0.05, would require only six participants to replicate the effect with a power of 95%. Although the design of that study differs from ours in several ways, it is not unreasonable to assume that if reformulations are robustly used for learning, a sample size that is already nine times larger than that reported by Saxton (1997) should be able to capture the effect. If not, we are forced to conclude that the effect size for any reformulation-based incidental learning must be much smaller than that reported by Saxton (1997).

3.1.5. Statistical analysis

Unless stated otherwise, all analyses in this paper were carried out using mixed effect models with the *lme4* package (Bates et al., 2015) in R (version 4.2.1, R Core Development Team, 2022). Following recommendations by Barr et al. (2013), we aimed for including the maximal random effect structure, and when the model did not converge, removed slopes until convergence was reached. Condition was center-coded as −0.5 for Reformulated and 0.5 for Modeled. For error analyses, correct responses were coded as 0 and an error was coded as 1. *p*-Values were computed through Satterthwaite approximation using the *LmerTest* package (version 3.1-3, Kuznetsova et al., 2017).

3.2. Results

Fig. 4 shows the pluralization error rates during each phase of the experiment. We first tested whether children in both the Modeled and Reformulated conditions overregularized the irregular nouns at similar rates at the pre-test. This model had pre-test error as its dependent variable and condition (Modeled vs. Reformulated) as its independent variable. Results showed comparable average error rates at pre-test across the two conditions ($M_{Modeled} = .68$, $M_{Reformulated} = .69$, $\beta_{Modeled} = -0.15$, $z_{Modeled} = -0.57$, $p = .57$). This finding is important, because it assures us that there were no significant baseline differences between the two groups in terms of their pluralization knowledge.

Next, we turn to learning. Average post-test error rates for the Modeled and Reformulated conditions were .48 and .64, respectively, suggesting a reduction of ∼20% in the Modeled and ∼ 5% in the Reformulated condition. To assess differences in learning between the two conditions, we predicted error rates from Phase (pre- vs. post-test), Condition (Modeled vs. Reformulated), and their interaction. The model only tolerated the random intercept of subjects and items, both of which were included. As shown in Table 3, children produced significantly fewer errors in the post-test compared to pre-test, showing evidence of learning ($z = -4.13$, $p < .001$). There were also marginally fewer errors in the Modeled condition ($z = -1.80$, $p = .071$). Importantly, learning was better in the Modeled compared to the Reformulated condition, as suggested by a significant interaction between Phase and Condition ($z = -2.09$, $p = .037$).

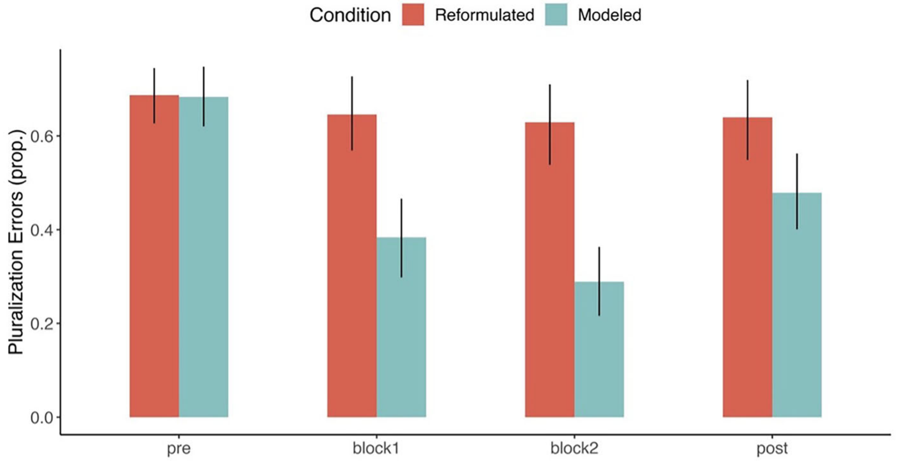

Fig. 4. Average proportions of pluralization errors by condition in the four phases of Experiment 1.
Note. Values show mean of subject means with 95% confidence intervals estimated by nonparametric bootstrapping. Note that in blocks 1 and 2, Reformulated and Modeled have had different exposures to input. The critical comparison is between pre- and post-test performance.

Table 3
Results of model predicting error rate from the interaction of phase of the experiment (pre or post) and condition for Experiment 1

Fixed effect	Coefficient	SE	z	p-value
Intercept	0.579	0.378	1.533	.125
Phase	−0.707	0.171	−4.133	< .001
Condition	−0.483	0.268	−1.803	.071
Phase * Condition	−0.710	0.340	−2.091	.037

To unpack the interaction observed in the model, we next tested pre- to post-test improvement individually in each condition. Such improvement was significant in the Modeled ($\beta_{Post} = -1.10$, $z = -4.70$, $p = < .001$) but not in the Reformulated ($\beta_{Post} = -0.34$, $z = -1.38$, $p = .168$) condition.

3.3. Discussion

Results of Experiment 1 showed that learning was significantly better in the Modeled than in the Reformulated condition. In fact, there was no robust evidence of learning in the Reformulated condition. This finding is not unexpected given the results of our corpus analysis, as well as past findings that also reported no clear evidence of immediate learning from reformulations (Morgan et al., 1995). However, they are diametrically opposed to the experimental findings reported by Saxton (1997) and Saxton and colleagues (1998), who demonstrated much better learning after reformulations than modeling. We maintain that the current design

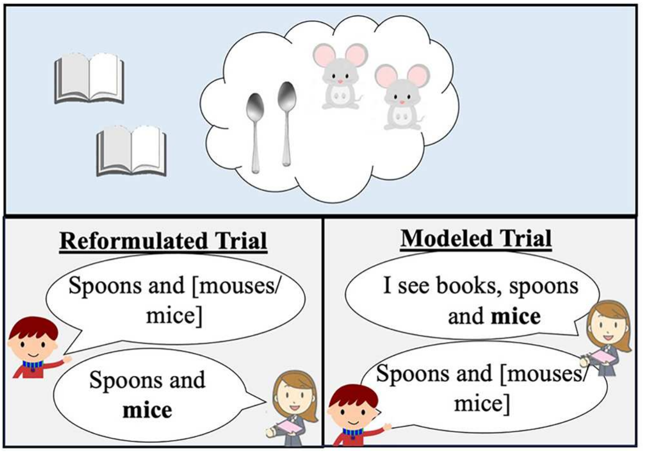

Fig. 5. An example of Experiment 2 training phase for Modeled and Reformulated trials.

is more representative of language learning in natural environments, where 3- to 4-year-old children are much more likely to learn words incidentally, as opposed to under explicit pedagogical instructions. That said, one could argue that the critical information in the current task was conveyed by the *numbers* and not the *nouns*. It is thus possible that the task goal distracted children from processing the nouns. Consequently, a correction on an erroneous noun was not picked up, as it was not directly relevant to the task goal.

Although this finding is informative in its own right, it is reasonable to assume that there are many situations in which the plural noun is more directly relevant to the task goal and, therefore, more prominently in children's focus of attention. If so, one could reasonably object that reformulations have not been given their best shot against modeling. Experiment 2 was designed to address this issue.

4. Experiment 2

Experiment 2 was designed to give incidental error-based learning its best shot in two ways. First, to make plural nouns more directly relevant to the task goal, we changed the design such that children were only presented with one dream cloud and a number of objects that moved into it. Their task was to simply tell the experimenter which objects moved into the cloud, for example, "spoons and mouses/mice" (Fig. 5). Similar to Experiment 1, in the Reformulated condition, the experimenter provided the correct utterance after the child. In the Modeled condition, the experimenter demonstrated the correct plural forms, for example, "I see trees, dogs, and children, but which go into my dream?" and the child responded by repeating the correct subset. Second, to increase statistical power, we used a within-subject design with an increased sample size. The switch from a between- to within-subject design also removes concerns about differences in the baseline learning abilities between children.

4.1. Method

4.1.1. Participants

Results of Experiment 1 suggested that even if children can learn through incidental processing of reformulations, the size of this effect would be much smaller than that reported in Saxton (1997). We, therefore, adopted a different approach to sample size estimation in Experiment 2. We hypothesized the smallest effect size of interest that may have any clinical significance (estimated at 0.3), and determined that with an alpha of 0.05, a sample size of 90 participants would allow us to detect an effect with a power of 0.8 in a within-subject design. To account for possible attrition, we preregistered the study with a sample size of 100 (Preregistration: https://osf.io/r6m9q?view_only=c923e0c88d954d61920009067dff8dc4). Data were collected from 102 3- and 4-year-olds over Zoom. Two children did not finish the game and were replaced to reach our preregistered total of 100 children. The final sample contained 100 children (46 biologically female children, 3.00−4.98 years, $M_{age} = 3.93$, $SD_{age} = 0.58$). Families were recruited in the same manner as Experiment 1, in addition to from local schools, and online platforms Children Helping Science, Reddit, and NextDoor. They were offered the same $10 gift card compensation. The study was approved by the university's IRB.

4.1.2. Materials

Children were tested on seven of the same irregular nouns (mice, geese, teeth, feet, men, women, snowmen). One item, "leaves," was exchanged for "children" because the plural forms ("leaves"/"leafs") were phonetically difficult to distinguish, causing the largest number of item-level exclusion in Experiment 1 due to irreconcilable disagreements between the two coders. The same eight regular plural nouns were used as fillers (snakes, frogs, pigs, cats, dogs, spoons, forks, turtles). However, the structure of this task added a third item that the experimenter named in the Modeled condition (see Fig. 5). Therefore, we wanted to balance exposure to the filler plural items that the child would produce along with the target items. We thus included another set of five filler plurals (cars, trees, books, shoes, and apples; see Appendix B for frequencies). Only the eight original regular plurals were named in conjunction with the irregular plurals by the child.

4.1.3. Procedure

As in Experiment 1, after informed consent was obtained by parents, children were told they would help fix the experimenter's dream machine. This game also took around 15 min to complete. The structure of the session was similar to Experiment 1, with a pre-test, two training blocks, and a post-test. Rather than immediately entering the pretest as in Experiment 1, children first labeled the images in their singular form to establish the root word. This initial introduction was included to avoid correcting children in the pre-test. Children then proceeded to the pre-test. Here, the items always appeared in pairs, but we asked children to say the number "two" along with the label to encourage the use of plural in this phase (e.g., "two mice").

During training blocks 1 and 2, three pairs of pictures appeared simultaneously outside the dream cloud, and then two of the three pairs moved into the cloud. The child was instructed to tell the experimenter which pictures went into the cloud (Fig. 5). In the Reformulated condition, the experimenter first asked the child "What was in my dream?". She then pressed a button to move the relevant objects into the dream cloud. The child then replied, for example, "Spoons and mice/mouses!". The experimenter then provided the correct response, for example, "Spoons and mice!". In the Modeled condition, the experimenter labeled all the pairs first, by saying "I see books, spoons, and mice, but which go into my dream?", and then pressed a button to move the relevant objects to the dream cloud. The child then replied with a subset of modeled utterances, for example, "Spoons and mice!". Next, the experimenter pressed a button to start a new trial.

Unlike Experiment 1, Experiment 2 used a within-subjects design in which each child completed trials from both conditions with different words. The assignment of words to Modeled and Reformulated conditions was counterbalanced across participants. Modeled and Reformulated trials were interleaved throughout the experiment, with two random orders of items counterbalanced across participants. Using PowerPoint's "presenter view," the experimenter had privileged access to each trial type and produced the prompt accordingly. Finally, at post-test, children named the pairs again, for example, "two mice," to assess whether there was any difference in learning between the words they repeated compared to the words the experimenter reformed.

4.1.4. Data preprocessing

Zoom sessions were recorded and transcribed by five coders, all of whom were native speakers of American English. The coders then marked whether the child made a lexical or plural error during each trial. To ensure reliable coding, 20 of the videos were coded independently in pairs, and coder judgments were reliably similar across error types (lexical average Cohen's $\kappa = 0.99$ and plural: $\kappa = 0.99$). Using the same criterion for exclusion as in Experiment 1 (> 20% lexical errors), one subject was excluded from the analysis. Therefore, the final sample included in the analysis is 99 children (45 biologically female, $M_{age} = 3.94$, $SD_{age} = 0.57$). As in Experiment 1, any trials in which the child used the wrong lexical item were removed from analyses, which was 5.2% of trials.

4.2. Results

Fig. 6 shows the results of Experiment 2. Statistical models were similar to those used in Experiment 1.

Pre- and post-test error rates for the Modeled condition were .81 and .55, respectively, showing an improvement of ∼26%. The same categories for the Reformulated condition were .85 and .61, suggesting an improvement of ∼ 24%. As in Experiment 1, to assess differences in learning between the two conditions, we predicted error rates from Phase (pre- vs. post-test), Condition (Modeled vs. Reformulated), and their interaction. The random effect structure was identical to Experiment 1. Results are shown in Table 4. Children again produced significantly fewer errors in the post-test compared to the pre-test, showing evidence of learning ($z = -7.55$, $p < .001$). This time, there was no significant effect of Condition or an interaction between Phase and Condition.

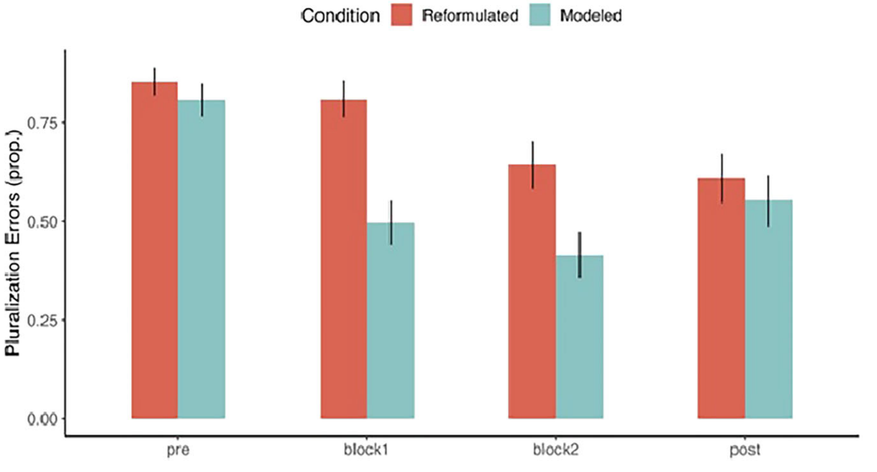

Fig. 6. Average proportions of pluralization errors by condition in the four phases of Experiment 2.
Note. Values show mean of subject means with 95% confidence intervals estimated by nonparametric bootstrapping. Note that in blocks 1 and 2, Reformulated and Modeled have had different exposures to input. The critical comparison is between pre- and post-test performance.

Table 4
Results of model predicting error rate from the interaction of phase of the experiment (pre or post) and condition for Experiment 2

Fixed effect	Coefficient	SE	z	p-value
Intercept	2.138	0.309	6.918	< .001
Phase	−1.572	0.208	−7.554	< .001
Condition	0.288	0.224	−1.286	.199
Phase * Condition	−0.002	0.282	−0.008	.994

The absence of a significant effect of Condition or a Phase by Condition interaction suggests comparable learning between the two conditions. To confirm that there was a significant improvement from pre- to post-test in both conditions, we looked at each condition separately. Improvement from the pre- to post-test was significant for both Modeled ($\beta_{post} = -1.60$, $z = -8.12$, $p = < .001$) and for Reformulated ($\beta_{post} = -1.65$, $z = -7.85$, $p = < .001$).

4.3. Discussion

Experiment 2 was designed to give error-based learning from reformulations its best shot, while still simulating an interactive play situation without an explicit pedagogical role to more closely capture the situations in which young children learn language. We replicated the benefit of errorless learning through modeling from Experiment 1. But this time, we also found that error-based learning through reformulations can benefit learning. However, even with increased power and target utterance being the focus of the communicative message, we found no evidence that error-based learning was superior to errorless learning as implied in pedagogical studies such as Saxton (1997) and Saxton et al. (1998).

5. General discussion

In a corpus analysis and two experiments, we tested whether the clear superiority of error-based learning over errorless learning observed in pedagogical settings extends to nonpedagogical settings in children younger than 5 years of age. We aimed to compare how these two learning mechanisms may contribute to children's learning of correct irregular forms. The corpus analysis found that adults reformulated children's errors a quarter of the time, comparable to Hirsh-Pasek and colleagues (1984) and somewhere between other past reports, ranging from 15% to 48% (Bohannon & Stanowicz, 1988; Chouinard & Clark, 2003; Farrar, 1992; Morgan et al., 1995). Additionally, we found that the IG for reformulation, while higher than replays, was modest (2% over a baseline of 92%). Finally, we did not find much evidence in the corpus that children made immediate use of reformulations to change their errors to correct utterances. In fact, in the few cases that children did reattempt to produce the target word, they were more likely to repeat their own errors rather replacing it with the correct form.

The results of Experiment 1 closely mirrored the findings of the corpus analysis. Repeating the correct form following the model from the experimenter was much more effective in changing initial pluralization errors to correct forms than processing reformulations. In fact, we found no robust evidence that reformulations benefitted learning when the target nouns were not the critical focus of the communicative message. Experiment 2 was designed to maximize the chance of observing a potential benefit of reformulations. In a within-subject design with increased power and better control for between-subject variability, we tested learning from modeling versus reformulations when the target nouns were the focus of children's communicative message. This time, we did observe a benefit of reformulations. However, even under these circumstances, there was no evidence that error-based learning was superior to errorless learning. In summary, the study provides converging evidence against error-based learning with incidental feedback as a more powerful mechanism (compared to errorless learning) for morpholexical learning in children of this age, at least over short periods of time.

5.1. *Why might error-based learning be difficult in children?*

The current findings may be surprising, given prior research on the efficacy of error-based learning and the centrality of this mechanism in many prominent models of learning in cognitive psychology. The results directly oppose those in similar designs in individuals with aphasia (Middleton et al., 2015; but see Fillingham et al., 2006) and even children (Saxton, 1997; Saxton et al., 1998). Additionally, these results may be seen as contradictory to a larger body of research on the testing effect, which shows that a retrieval attempt, even if it leads to an error, leads to better learning than simply studying the materials again (e.g., Bjork, 1988; Karpicke & Blunt, 2011; Rawson & Dunlosky, 2011; see Roediger & Butler, 2011, and Metcalfe, 2017, for reviews). Why might error-based learning not be the most potent mechanism for learning in children? There are a few possible reasons for this. First, children make many more errors than adults and past research has shown that they detect/correct a lower proportion of those errors, even when they do possess the knowledge to switch to the correct form (Hanley et al., 2016; Nozari & Omaki, 2022). This prior finding indicates poorer self-monitoring abilities in children, which can explain the difference in the efficacy of error-based learning mechanisms in pedagogical settings, where the explicit external feedback reduces the need for self-monitoring to guide learning, compared to nonpedagogical settings, where self-monitoring is the key for initiating learning. In keeping with this idea, when the relevant information (i.e., the plural noun) was made the focus of the communicative message, that is, when monitoring the relevant nouns became necessary for accomplishing the task goal, children started to benefit from reformulations.

A second possible reason for the difficulty of error-based learning is that child language production is marked by more preservatory behavior, compared to adult language (Stemberger, 1989; see Hepner & Nozari, 2020, for a similar demonstration in other production systems with decreased signal to noise ratio). Both adults and children produce speech errors by anticipating or perseverating words and segments. For example, the target sentence "Put the book on the table" could turn into "Put the *table* on the table" (anticipation) or "Put the book on the *book*" (perseveration). Similarly, the word "put" may be mistakenly pronounced as /t̲ʊt/ (anticipation) or /pʊp̲/ (perseveration). The ratio of anticipations to perseverations is considerably higher in children than in adult speakers, indicating that they have trouble suppressing what they have already produced (Dell et al., 1997; Stemberger, 1989). In the current context, it is easier to learn by modeling than by reformulation, because the child has produced the incorrect form more times in the Reformulated condition, which compounds the problem of suppressing the past. In keeping with this possibility, children in our corpus analysis repeated more of their own errors than caregivers' corrections after reformulations.

The point about the difficulty of recovering from errors is worth emphasizing. While previous research has demonstrated how children make powerful statistical inferences based on parent input (Saffran & Wilson, 2003; Scott & Fisher, 2012; Smith & Yu, 2008), children's own production is not always considered as a critical factor in models of language acquisition (c.f., Tomasello, 2000; Clark, 2020). This is quite surprising, given the wide range of variables covered by such models, such as the quality of parent speech, socioeconomic status, the focus of child's spatial attention and gaze, sensitivity to informativeness, and capacity of working memory (e.g., Bakopoulou et al., 2023; Bhat et al., 2022; Bohn et al., 2021; Dailey & Bergelson, 2022). The findings of the current study lend further support to theories of language learning such as usage-based theory and dynamic systems theory, which emphasize that language use, such as production, is an important factor in shaping the learning environment and outcomes (Tomasello, 2000; Van Geert & Verspoor, 2015). In the same vein, there are now reports that producing language even affects how well adults learn to *understand* language (Hopman & MacDonald, 2018; cf., Waller et al., 2024), not to mention that a few production attempts are enough to implicitly change long-standing production patterns, such as phonotactic or orthotactic constraints (Atilgan et al., 2022; Dell et al., 2000; Smalle et al, 2017; Warker & Dell, 2006; Warker et al., 2009). In short, the current findings strongly suggest that language learning is not only influenced by external speech input, but also by the learner's own production and call for the better integration of production into models of language acquisition in childhood.

5.2. *When is error-based learning effective?*

Although the current results did not support error-based learning as a particularly powerful mechanism in language learning in children, it does not mean that such a mechanism cannot be effective. Rather, error-based learning seems to be effective when certain conditions are met. One such condition is when the task goal is specifically word learning, that is, when the child's attention is explicitly directed toward feedback as a critical signal for successfully completing the task (Saxton, 1997; Saxton et al., 1998). Interestingly, our results suggest that task goal may be graded. Even when the goal was not explicitly to learn words, children in Experiment 2 did show learning in the error-based condition when the to-be-learned items were the focus of discourse and directly relevant to the communicative goal.

Error-based learning may also be more effective for learning new words than for correcting mistakes on old ones, because the errors produced during novel word learning are usually less-practiced than those produced on already-known words. For example, in Saxton's studies, children have probably never produced the error "pelled" before, since its stem does not exist in the lexicon. Whereas children in our study may have produced "mouses" and similar words many times in the past. In the latter case, committing the error once again would reinforce an existing connection that would be harder to undo through error-based learning.

Another favorable condition for error-based learning is when predictions exist in the child's language system *without resulting in overt errors* (Fazekas et al., 2020; Lin & Fisher, 2017; Peter et al., 2015). It appears that the learning of the error, that is, strengthening the link between the error and the concept through Hebbian learning, which follows overt error production, can be powerful enough to override the benefit of error-based learning from internal error signal. In caregiver-child play time, this can be incorporated by caregivers taking a more active role in modeling names, for example, naming picture in a picture book, for the child, rather than asking them to try out names. In pedagogical settings, this can be remedied by encouraging the child to repeat the correct form after feedback has been provided, to undo the undesirable learning of the erroneous form. We did not instruct children to do so in our experiment, because the corpus results suggested that children rarely spontaneously repeated after a reformulation, and our goal was to assess the efficacy of error-based learning under circumstances that were better reflective of natural language learning environments. But it is quite possible that asking the child to repeat the correct form after reformulations, as is often practiced in pedagogical settings, would have led to better error-based learning (see Middleton et al., 2015 for the success of this method in individuals with aphasia).

5.3. *Are errorless and error-based mechanisms completely separable?*

A potential objection to the design of all errorless versus error-based learning studies reviewed in this paper (Middleton et al., 2015; Saxton, 1997; Saxton et al., 1998; Schuchard

& Middleton, 2018) could be that the Modeled condition is not entirely errorless. If children generate the incorrect prediction internally, the correct form uttered by the experimenter could generate an error signal that drives error-based learning. This is similar to the mechanism assumed in studies of syntactic priming/learning with matched or mismatched verb bias reviewed in the introduction (Fazekas et al., 2020; Lin & Fisher, 2017; Peter et al., 2015). In other words, what is assumed here is error-based learning through perception which transfers to production. Although this is a possibility, it is not highly likely in the context of the current experiment. Pickering and Gambi (2018) present a distinction between two mechanisms of predictive processing that is helpful: they claim that associative prediction (e.g., activating "key" when you hear "lock") is automatic (e.g., Kamide et al., 2003), whereas the kind of activation that transfers from perception to production requires deliberate activation of the relevant representations in the production system, an action that is much less common than the first type of prediction. They further argue that the latter is even less common in children (e.g., Borovsky et al., 2014). In our own task, there is little time between the presentation of the pictures that contain multiple entities and the utterance in the Modeled condition, leaving little room for prediction before hearing the correct response.

Thus, the goal is not to claim that *no* error signal—and consequently no error-based learning—is present in the Modeled condition. Similarly, even in the Reformulated condition, one could not rule out that there is some Hebbian learning at work, binding feedback to the picture at the end of the process. Rather, the key assumption of the current study is that the relative weight of the two learning mechanisms is different in the two conditions. We maintain that this assumption is correct: the Modeled condition taps much more strongly into errorless learning by preventing consistent prediction and the full instantiation of such prediction through the production system. The Reformulated condition, on the other hand, strongly taps into error-based learning mechanisms.

5.4. Limitations and future directions

There are a few other possibilities that must be mentioned as potential limitations of the current study. One caveat is the limited number of exposures to each word. We intentionally kept the study time to under 20 min to maintain the attention of children as young as 3 years of age, which naturally limited the amount of exposure per target. It is possible that more exposures could have led to learning from reformulations even in Experiment 1. However, the fact that in Experiment 2 children showed evidence of learning from reformulations only after a single exposure (i.e., as soon as block 2) implies that error-based learning after limited exposure is possible when other conditions are favorable to such learning.

Another caveat is that we did not test learning and retention over longer periods of time. Indeed, some studies suggest that the effects of negative evidence, such as reformulations, greatly increase over time in pedagogical settings (Saxton et al., 1998). We cannot rule out a similar possibility in nonpedagogical settings. Modeling may help learning in the short term, but this learning may not be as robust over time as receiving corrections. Finally, it is possible that a combination of factors, such as number of reformulations, the interval between subsequent reformulations, and the timing of the test may alter the pattern of results observed in the current study, as such factors have been shown to influence learning and retention, at least in pedagogical settings (e.g., Landauer & Bjork, 1978; see Roediger & Butler, 2011, for a review). Investigating the effect of these factors and their modulatory effect on incidental errorless versus error-based learning in children is a great avenue for future research.

We also acknowledge that there may be differences in learning different parts of speech. For example, the studies of Saxton (1997) and Saxton et al. (1998) tested the effects of reformulation on past tense rather than pluralization. While Chouinard and Clark (2003) found reformulation rates across various errors (lexical, morphological, syntactic, and phonological) to be relatively similar, it is unclear whether children notice these corrections and learn from them in each at the same rate. There are clear differences in learning phonological, lexical, and morpho-syntactic aspects of a language, and the current study was not designed to compare learning across these aspects. No single experimental design examining young children can include all facets of language learning while ensuring enough power for statistical analysis. Therefore, we view this study as examining one facet, which provides a basis for comparison with other facets in future studies.

6. Conclusion

The results of the current study provided converging evidence that error-based learning is not a particularly powerful mechanism for immediate morpholexical learning in nonpedagogical settings in 3- and 4-year-old children, at least when compared to errorless learning. In contrast, error-based learning seems to be effective when predictions are formed in the child's language system without surfacing as overt errors or in pedagogical settings where the importance of feedback in correcting errors is made explicit to the child.

Acknowledgments

We thank Angela Wei, Riya Chaudhry, Lily Madojemu, and Christina Choi for their work transcribing and coding videos for this project, as well as all the families who participated.

Funding

This work was supported by the James S. McDonnell Foundation Scholar Award in Understanding Human Cognition #220020506 to DY and NSF BCS-2317121 to NN.

Open Research Badges

This article has earned Open Data and Open Materials badges. Data and materials are available at https://osf.io/bw8zh/?view_only=c923e0c88d954d61920009067dff8dc4.

References

Arnon, I., & Clark, E. V. (2011). Why brush your teeth is better than teeth–Children's word production is facilitated in familiar sentence-frames. *Language Learning and Development*, *7*(2), 107–129.

Atilgan, N., Chu, L., & Nozari, N. (2022). Statistical learning in language production: Studying orthotactic constraints through typing. In *Oral presentation at the 63rd Annual Meeting of the Psychonomic Society*. Boston, MA.

Baese-Berk, M. M., & Samuel, A. G. (2016). Listeners beware: Speech production may be bad for learning speech sounds. *Journal of Memory and Language*, *89*, 23–36.

Bakopoulou, M., Lorenz, M. G., Forbes, S. H., Tremlin, R., Bates, J., & Samuelson, L. K. (2023). Vocabulary and automatic attention: The relation between novel words and gaze dynamics in noun generalization. *Developmental Science*, 26(6), e13399. https://doi.org/10.1111/desc.13399

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Behrens, H. (2021). Constructivist approaches to first language acquisition. *Journal of Child Language*, *48*(5), 959–983.

Bhat, A. A., Spencer, J. P., & Samuelson, L. K. (2022). Word-object learning via visual exploration in space (WOLVES): A neural process model of cross-situational word learning. *Psychological Review*, *129*(4), 640–695.

Biedermann, B., Beyersmann, E., Mason, C., & Nickels, L. (2013). Does plural dominance play a role in spoken picture naming? A comparison of unimpaired and impaired speakers. *Journal of Neurolinguistics*, *26*(6), 712–736.

Bjork, R. A. (1988). Retrieval practice and the maintenance of knowledge. In M. M. Gruneber, P. E. Morris & R. N. Sykes (Eds.), *Practical aspects of memory: Current research and issues* (Vol. 1, pp. 396–401).

Bohannon, J. N., & Stanowicz, L. B. (1988). The issue of negative evidence: Adult responses to children's language errors. *Developmental Psychology*, *24*(5), 684–689.

Bohn, M., Tessler, M. H., Merrick, M., & Frank, M. C. (2021). How young children integrate information sources to infer the meaning of words. *Nature Human Behaviour*, *5*(8), 1046–1054.

Borovsky, A., Sweeney, K., Elman, J. L., & Fernald, A. (2014). Real-time interpretation of novel events across childhood. *Journal of Memory and Language*, *73*, 1–14.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977–990.

Cartmill, E. A., Armstrong III, B. F., Gleitman, L. R., Goldin-Meadow, S., Medina, T. N., & Trueswell, J. C. (2013). Quality of early parent input predicts child vocabulary 3 years later. *Proceedings of the National Academy of Sciences*, *110*(28), 11278–11283.

Chouinard, M. M., & Clark, E. V. (2003). Adult reformulations of child errors as negative evidence. *Journal of Child Language*, *30*(3), 637–670.

Clark, E. V. (2020). Conversational repair and the acquisition of language. *Discourse Processes*, *57*(5–6), 441–459.

Dailey, S., & Bergelson, E. (2022). Language input to infants of different socioeconomic statuses: A quantitative meta-analysis. *Developmental Science*, *25*(3), e13192.

Dell, G. S., Burger, L. K., & Svec, W. R. (1997). Language production and serial order: A functional analysis and a model. *Psychological Review*, *104*(1), 123–147.

Dell, G. S., Reed, K. D., Adams, D. R., & Meyer, A. S. (2000). Speech errors, phonotactic constraints, and implicit learning: A study of the role of experience in language production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(6), 1355–1367.

Farrar, M. J. (1992). Negative evidence and grammatical morpheme acquisition. *Developmental Psychology*, *28*(1), 90.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149–1160.

Fazekas, J., Jessop, A., Pine, J., & Rowland, C. (2020). Do children learn from their prediction mistakes? A registered report evaluating error-based theories of language acquisition. *Royal Society Open Science*, *7*(11), 180877.

Fillingham, J. K., Hodgson, C., Sage, K., & Lambon Ralph, M. A. (2003). The application of errorless learning to aphasic disorders: A review of theory and practice. *Neuropsychological Rehabilitation*, *13*(3), 337–363.

Fillingham, J. K., Sage, K., & Lambon Ralph, M. A. (2006). The treatment of anomia using errorless learning. *Neuropsychological Rehabilitation*, *16*(2), 129–154.

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, *1*(1), 3–55.

Gleitman, L. R., Cassidy, K., Nappa, R., Papafragou, A., & Trueswell, J. C. (2005). Hard words. *Language Learning and Development*, *1*(1), 23–64.

Goldstein, M. H., & Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological Science*, *19*(5), 515–523.

Hanley, R.J., Cortis, C., Budd, M.J., & Nozari, N. (2016). Did I say dog or cat? A study of semantic error detection and correction in children. *Journal of Experimental Child Psychology*, *142*, 36–47.

Hebb, D. O. (1949). *The organization of behavior; A neuropsychological theory*. Wiley.

Hepner, C. R., & Nozari, N. (2020). The dual origin of lexical perseverations in aphasia: Residual activation and incremental learning. *Neuropsychologia*, *147*, 107603. https://doi.org/10.1016/j.neuropsychologia.2020.107603

Hirsh-Pasek, K., Treiman, R., & Schneiderman, M. (1984). Brown & Hanlon revisited: Mothers' sensitivity to ungrammatical forms. *Journal of Child Language*, *11*(1), 81–88.

Hopman, E. W., & MacDonald, M. C. (2018). Production practice during language learning improves comprehension. *Psychological Science*, *29*(6), 961–971.

Humphreys, K. R., Menzies, H., & Lake, J. K. (2010). Repeated speech errors: Evidence for learning. *Cognition*, *117*(2), 151–165.

Kamide, Y., Altmann, G. T., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, *49*(1), 133–156.

Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, *331*(6018), 772–775.

Keren-Portnoy, T., Vihman, M. M., DePaolis, R. A., Whitaker, C. J., & Williams, N. M. (2010). The role of vocal practice in constructing phonological working memory. https://doi.org/10.1044/1092-4388(2009/09-0003)

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*, 1–26.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, *44*(4), 978–990.

Landauer, T. K., & Bjork, R. A. (1978). Optimum rehearsal patterns and name learning. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes. (Eds.), *Practical aspects of memory* (pp. 625–632). Academic Press.

Lin, Y., & Fisher, C. (2017). Error-based learning: A mechanism for linking verbs to syntax. In *Proceedings of the 39th Annual Conference of the Cognitive Science Society*. Cognitive Science Society.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk. Transcription format and programs* (Vol. 1). Psychology Press.

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), i–178. https://doi.org/10.2307/1166115

Masek, L. R., McMillan, B. T., Paterson, S. J., Tamis-LeMonda, C. S., Golinkoff, R. M., & Hirsh-Pasek, K. (2021). Where language meets attention: How contingent interactions promote learning. *Developmental Review*, *60*, 100961.

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, *20*(2), 121–157.

McClelland, J. L., Rumelhart, D. E., & PDP Research Group. (1987). *Parallel distributed processing, volume 2: Explorations in the microstructure of cognition: Psychological and biological models* (Vol. 2). MIT Press.

Metcalfe, J. (2017). Learning from errors. *Annual Review of Psychology*, *68*, 465–489.

Middleton, E. L., & Schwartz, M. F. (2012). Errorless learning in cognitive rehabilitation: A critical review. *Neuropsychological Rehabilitation*, *22*(2), 138–168.

Middleton, E. L., Schwartz, M. F., Rawson, K. A., & Garvey, K. (2015). Test-enhanced learning versus errorless learning in aphasia rehabilitation: Testing competing psychological principles. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(4), 1253–1261.

Morgan, J. L., Bonamo, K. M., & Travis, L. L. (1995). Negative evidence on negative evidence. *Developmental Psychology*, *31*(2), 180–197.

Morgan, J. L., & Travis, L. L. (1989). Limits on negative information in language input. *Journal of Child Language*, *16*(3), 531–552.

Nappa, R., Wessel, A., McEldoon, K. L., Gleitman, L. R., & Trueswell, J. C. (2009). Use of speaker's gaze and syntax in verb learning. *Language Learning and Development*, *5*(4), 203–234.

Nelson, K. E., Carskaddon, G., & Bonvillian, J. D. (1973). Syntax acquisition: Impact of experimental variation in adult verbal interaction with the child. *Child Development*, *44*(3), 497–504.

Nozari, N. (2018). How special is language production? Perspectives from monitoring and control. In K. Federmeier & D. Watson (Eds.), *Psychology of learning and motivation: Current topics in language* (Vol. 68, pp. 179–213). Cambridge, MA: Academic Press.

Nozari, N., & Omaki, A. (2022). Revisiting agreement: Do children and adults compute subject-verb agreement differently? In J. Culbertson, A. Perfors, H. Rabagliati, & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society* (pp. 409–415). Cognitive Science Society.

Oppenheim, G. M., & Nozari, N. (2024). Similarity-induced interference or facilitation in language production reflects representation, not selection. *Cognition*, *245*, 105720.

Peter, M., Chang, F., Pine, J. M., Blything, R., & Rowland, C. F. (2015). When and how do children develop knowledge of verb argument structure? Evidence from verb bias effects in a structural priming task. *Journal of Memory and Language*, *81*, 1–15.

Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin*, *144*(10), 1002–1044.

Pinker, S. (1979). Formal models of language learning. *Cognition*, *7*(3), 217–283.

R Core Team. (2022). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. URL https://www.R-project.org/.

Ramscar, M., Dye, M., & McCauley, S. M. (2013). Error and expectation in language learning: The curious absence of " mouses" in adult speech. *Language*, *89*(4) 760–793.

Rawson, K. A., & Dunlosky, J. (2011). Optimizing schedules of retrieval practice for durable and efficient learning: How much is enough? *Journal of Experimental Psychology: General*, *140*(3), 283–302.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Proskay (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–66). New York: Appleton-Century-Crofts.

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, *15*(1), 20–27.

Roseberry, S., Hirsh-Pasek, K., & Golinkoff, R. M. (2014). Skype me! Socially contingent interactions help toddlers learn language. *Child Development*, *85*(3), 956–970.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1985). *Learning internal representations by error propagation*. California University San Diego La Jolla Institute for Cognitive Science.

Saffran, J. R., & Wilson, D. P. (2003). From syllables to syntax: Multilevel statistical learning by 12-month-old infants. *Infancy*, *4*(2), 273–284.

Saxton, M. (1997). The contrast theory of negative input. *Journal of Child Language*, *24*(1), 139–161.

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

Saxton, M., Kulcsar, B., Marshall, G., & Rupra, M. (1998). Longer-term effects of corrective input: An experimental approach. *Journal of Child Language*, *25*(3), 701–721.
Schuchard, J., & Middleton, E. L. (2018). The roles of retrieval practice versus errorless learning in strengthening lexical access in aphasia. *Journal of Speech, Language, and Hearing Research*, *61*(7), 1700–1717.
Scott, R. M., & Fisher, C. (2012). 2.5-year-olds use cross-situational consistency to learn verbs under referential uncertainty. *Cognition*, *122*(2), 163–180.
Smalle, E. H., Muylle, M., Szmalec, A., & Duyck, W. (2017). The different time course of phonotactic constraint learning in children and adults: Evidence from speech errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(11), 1821–1827.
Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558–1568.
Stemberger, J. P. (1989). Speech errors in early child language production. *Journal of Memory and Language*, *28*(2), 164–188.
Tomasello, M. (2000). First steps toward a usage-based theory of language acquisition. *Cognitive Linguistics*, *11*(1), 261–282.
Van Geert, P., & Verspoor, M. (2015). Dynamic systems and language development. In B. MacWhinney & William O'Grady (Eds.), *The handbook of language emergence* (pp. 537–555).
Waller, M., Yurovsky, D., & Nozari, N. (2024). Effect of similarity and training experiences on new vocabulary learning. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 46).
Warker, J. A., & Dell, G. S. (2006). Speech errors reflect newly learned phonotactic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(2), 387–398.
Warker, J. A., Xu, Y., Dell, G. S., & Fisher, C. (2009). Speech errors reflect the phonotactic constraints in recently spoken syllables, but not in recently heard syllables. *Cognition*, *112*(1), 81–96.
Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review*, *114*(2), 245–272.

Appendix A

Walkthrough of Information Gain Analyses:
Information Gain Calculation for Replays:

$$IG\left(Accuracy\left(Acc\right), Replay\left(Rep\right)\right) = Entropy\left(Acc\right) - Entropy\left(Acc, Rep\right)$$

$$Entropy\left(X\right) = \sum_{i=1}^{c} -p_i \log_2 p_i$$

$$Entropy\left(Acc\right) = -\left(\frac{Total\ Correct}{Total\ Utterances} * \log_2 \frac{Total\ Correct}{Total\ Utterances} + \frac{Total\ Errors}{Total\ Utterances} * \log_2 \frac{Total\ Errors}{Total\ Utterances}\right)$$

$$Entropy\left(Acc\right) = -(0.919 * \log_2 0.919 + 0.0815 * \log_2 0.0815) = 0.407$$

$$Entropy\left(X, Y\right) = \sum_{c \in Y} P\left(c\right) E\left(c\right)$$

$$Entropy\left(Acc, Rep\right) = p\left(Rep_{Yes}\right) * Entropy\left(Rep_{Yes}\right) + p\left(Rep_{No}\right) * Entropy\left(Rep_{No}\right)$$

$$Entropy\,(Acc, Rep) = \frac{Total\ Replays}{Total\ Utterances} * \left(\frac{Total\ Replays\ to\ Correct\ Utts}{Total\ Replays} * \log_2 \frac{Total\ Replays\ to\ Correct\ Utts}{Total\ Replays} + \frac{Total\ Replays\ to\ Errors}{Total\ Replays} * \log_2 \frac{Total\ Replays\ to\ Errors}{Total\ Replays}\right) + \frac{Total\ Not\ Replay\ Utts}{Total\ Utterances} * \left(\frac{Total\ Not\ Replay\ to\ Correct\ Utts}{Total\ Not\ Replays} * \log_2 \frac{Total\ Not\ Replay\ to\ Correct\ Utts}{Total\ Not\ Replays} + \frac{Total\ Not\ Replay\ to\ Errors}{Total\ Not\ Replays} * \log_2 \frac{Total\ Not\ Replay\ to\ Errors}{Total\ Not\ Replays}\right)$$

$$Etropy\,(Acc, Rep) = 0.367 * (0.964 * \log_2 0.964 + 0.0361 * \log_2 0.0361) + 0.633 * (0.892 * \log_2 0.892 + 0.108 * \log_2 0.108) = 0.394$$

$$IG\,(Accuracy\,(Acc)\,,Replay\,(Rep)) = 0.407 - 0.394 = 0.013$$

Reformulation Information Gain Calculation:

$$IG\,(Accuracy\,(Acc)\,, Reformulation\,(Ref)) = Entropy\,(Acc) - Entropy\,(Acc, Ref)$$

$$Entropy\,(X) = \sum_{i=1}^{c} -p_i \log_2 p_i$$

$$Entropy\,(Acc) = -\left(\frac{Total\ Correct}{Total\ Utterances} * \log_2 \frac{Total\ Correct}{Total\ Utterances} + \frac{Total\ Errors}{Total\ Utterances} * \log_2 \frac{Total\ Errors}{Total\ Utterances}\right)$$

$$Entropy\,(Acc) = -(0.919 * \log_2 0.919 + 0.0815 * \log_2 0.0815) = 0.407$$

$$Entropy\,(X, Y) = \sum_{c \in Y} P\,(c)\,E\,(c)$$

$$Entropy\,(Acc, Ref) = p\left(Ref_{Yes}\right) * Entropy\left(Ref_{Yes}\right) + p\left(Ref_{No}\right) * Entropy\left(Ref_{No}\right)$$

$$Entropy\,(Acc, Ref) = \frac{Total\ Reforms}{Total\ Utterances} * \left(\frac{Total\ Reforms\ to\ Correct\ Utts}{Total\ Reforms} * \log_2 \frac{Total\ Reforms\ to\ Correct\ Utts}{Total\ Reforms} + \frac{Total\ Reforms\ to\ Errors}{Total\ Reforms} * \log_2 \frac{Total\ Reforms\ to\ Errors}{Total\ Reforms}\right) + \frac{Total\ Not\ Reform\ Utts}{Total\ Utterances} * \left(\frac{Total\ Not\ Reforms\ to\ Correct\ Utts}{Total\ Not\ Reforms} * \log_2 \frac{Total\ Not\ Reforms\ to\ Correct\ Utts}{Total\ Not\ Reforms} + \frac{Total\ Not\ Reforms\ to\ Errors}{Total\ Not\ Reforms} * \log_2 \frac{Total\ Not\ Reforms\ to\ Errors}{Total\ Not\ Reforms}\right)$$

$$Entropy\,(Acc, Ref) = 0.022 * (0.057 * \log_2 0.056 + 0.943 * \log_2 0.943) + 0.978 * (0.938 * \log_2 0.938 + 0.062 * \log_2 0.062) = 0.335$$

$$IG\,(Accuracy\,(Acc)\,, Replay\,(Rep)) = 0.407 - 0.335 = 0.071$$

Appendix B

Table B1
Log10 word frequency (per one million words) and age of acquisition (AoA) of the nontarget nouns in Experiments 1 and 2

Word	Singular frequency	Plural frequency	Mean AoA	SD AoA
snake	3.05	2.72	5.10	2.47
frog	2.78	2.23	4.32	1.45
pig	3.30	2.83	3.84	1.00
cat	3.53	3.00	3.68	1.52
dog	3.99	3.43	2.80	1.20
spoon	2.59	2.01	2.50	0.79
fork	2.65	2.06	3.63	1.86
turtle	2.93	2.18	4.17	1.58
book	3.95	3.53	3.68	1.80
tree	3.52	3.19	3.57	1.61
apple	3.08	2.62	4.15	1.81
shoe	3.19	3.58	2.60	1.19
car	4.39	3.37	3.37	1.95