# Error and expectation in language learning: An inquiry into the many curious incidents of "mouses" in adult speech

**Michael Ramscar & Melody Dye**
Department of Psychology, Stanford University,
Jordan Hall, Stanford, CA 94305.

## Abstract

Much can be learned about the world by examining the discrepancies between what is expected and what actually occurs. Although many formal learning theories make use of prediction error as an important— even necessary—component in explaining behavior, this source of evidence has been largely overlooked in the language-learning literature. In this paper, we show how incorporating prediction error into a model of plural word learning (Ramscar & Yarlett, 2007) can yield a surprising prediction: that at an appropriate point in learning, the tendency of children to over-regularize irregular plurals can be reduced, by exposing them to regular plurals alone. We report on an experiment, which was designed to test the model's predictions empirically. The findings indicate that memory testing on regular plurals led to significant reductions in the rates of plural over-regularization in six-year-olds.

## Introduction

*Gregory: "Is there any other point to which you would wish to draw my attention?"*
*Holmes: "To the curious incident of the dog in the night-time."*
*Gregory: "The dog did nothing in the night-time."*
*Holmes: "That was the curious incident."*
"Silver Blaze," Sir Arthur Conan Doyle.

A racehorse vanishes on the eve of an important race, its trainer murdered. Sherlock Holmes lights upon a crucial piece of evidence: a dog on the premises has remained silent throughout the time in question. The fact that the dog did not bark – and thus, that an expected event did not occur – proves an important clue to the identity of the murderer. As the curious incident of the dog in the nighttime reminds us, much can be learned from discrepancies between what is expected and what actually occurs.

In what follows, we show how in the ordinary course of their lives, people use the discrepancy between what they expect and what they actually experience as a vital source of information in learning; and that often, as in the case of Sherlock Holmes and The Silver Blaze, the non-occurrence of expected events provides important negative evidence. That people use such evidence is only natural: expectation and prediction-error are important components of animal learning (Rescorla, 1988). However, these factors have been largely overlooked in discussions of children's learning, especially in relation to language. The extensive literature asserting the lack of negative evidence to children learning language (e.g., Chomsky, 1959; Pinker, 1984, 2004; Marcus, 1993) either ignores expectation and error-driven learning, or treats them superficially at best. Expectation is usually dismissed as a weak form of 'indirect negative evidence' that can offer little to no assistance in the complex process of language acquisition (Pinker, 2004). Here we show that prediction-error provides an abundant source of evidence in human learning, and in particular language learning, by testing and confirming an intriguing prediction that error-driven learning makes about children's plural over-regularization errors: namely, that at an appropriate point in learning, the tendency of children to over-regularize irregular plurals can be reduced through exposure to regular plurals alone.

## Prediction error and learning theory

Formal learning models are able to account for a wide range of the effects associated with learning by assuming that learning is driven by the discrepancy between what is expected and what is actually observed (error-driven learning). The learned predictive value of cues produces expectations, and any difference in the value of what is expected versus what is experienced produces further learning. In the Rescorla-Wagner (1972) model, for example, the change in associative strength between a stimulus i and a response (or event) j on trial n is defined as: [1]

$$\Delta V_{ij}^{n} = \boldsymbol{\alpha}_i \boldsymbol{\beta}_j \, (\boldsymbol{\lambda}_j - V_{total}) \qquad (1)$$

Learning is governed by the value of $(\lambda_j - V_{TOTAL})$ where $\lambda_j$ is the value of the predicted event and $V_{total}$ is the predictive value of a set of cues. In the ordinary course of learning, the discrepancy between $\lambda_j$ and $V_{total}$ reduces over repeated trials, producing a negatively accelerated learning curve, and asymptotic learning.

What is often overlooked is what happens when a predicted event does not occur. If a cue predicts something that doesn't follow, then $\lambda_j$ will have a value

[1] $n$ indexes the current trial. $0 \leq \alpha_i \leq 1$ denotes the saliency of cue $_i$, $0 \leq \beta_j \leq 1$ denotes the learning rate of event$_j$, $\lambda_j$ denotes the maximum amount of associative strength that cue$_j$ can support, and $V_{total}$ is the sum of the associative strengths between all cues$_j$ present on the current trial and event$_j$.

of zero for that trial. In this case the discrepancy ($\lambda_j$ - $V_{TOTAL}$) will have a negative value, resulting in a reduction in the associative strength between the cues present on that trial and the absent feature *j*. For example, in modeling learning in a dog being trained to expect food when a bell is sounded, setting $\lambda_j$ to 1 for training trials where food is given, and 0 for later trials when no food appears, allows for the characteristic patterns of training and extinction to be modelled. This means that latent learning about the relationship between cues and events that are not actually present occurs in these circumstances, and it is this process that is a key aspect of learning.

Thus, in error-driven learning, cues compete with one another for relevance, producing associative learning patterns that can differ greatly from those that would arise out of a record of the correlation between cues and outcomes (Rescorla, 1988). There is evidence for this mechanism at a neural level. Increases and decreases in the firing rates of monkeys' striatal dopamine neurons appear to track the degree to which the outcomes of training trials are under- or over-predicted (Hollerman & Schulz, 1998).

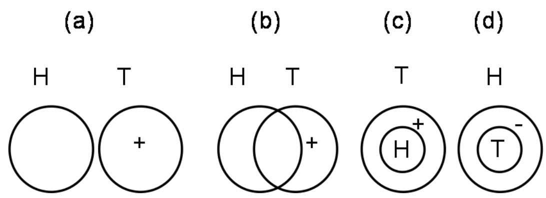


Figure 1. Four logical situations a child might arrive at while trying to "learn" a language (for the purposes of the example, language learning is assumed to be a process in which the child guesses the grammar that underlies that adult target language). Each circle represents the set of sentences constituting a language. "H" stands for the child's "hypothesized language"; "T" stands for the adult "target language." "+" indicates a grammatical sentence in the language the child is trying to learn, and "-" represents an ungrammatical sentence (Pinker, 1989).

### Expectation in language learning

A good example of the considerations that have led to the widespread belief that much of the conceptual structure of language is innate (see e.g. Pinker, 1984) is the "logical problem of language acquisition" (LPLA). A classic statement of this is provided by Pinker (1984) and is depicted in Figure 1. According to the LPLA, in attempting, to learn language, children "hypothesize the grammar of the adult language" (strictly, the child's task is to guess guessing the set of grammatical sentences that comprise a language; Gold, 1967).

Possible languages are depicted as circles corresponding to sets of word sequences, and four logical possibilities for how a child's hypothesis might differ from adult language are given. In the first possibility (a), the child's hypothesis language, H, is disjoint from the language to be acquired (the "target language" - T). In terms of noun usage, on which we focus here, this corresponds to the state of a child learning English who cannot produce any well-formed noun plurals (the child might say things like "the mouses" but never "the mice."). In (b), the sets H and T intersect, corresponding to a child who has learned some nouns correctly but others incorrectly (the child uses nouns like "mice" alongside incorrect words like "gooses"). In (c), H is a subset of T, which means that the child has mastered usage of some but not all English noun plurals and never uses forms that are not part of English. Finally, in (d), H is a superset of T, meaning that the child has mastered all English nouns but nevertheless produces some forms that are not part of the English language (i.e., the child says both "mouses" and "mice" interchangeably).

Since the LPLA assumes that learners cannot recover from erroneous inferences without corrective feedback, and because children do not get the kind of feedback required (Brown & Hanlon, 1970), in addition to the fact that they through stage (d), it follows accordingly that, children cannot acquire language simply by attending to the input. (Indeed, the idea that language is learned purely from experience is often regarded as having been effectively disproved; see Baker, 1979; Gold, 1967; Pinker, 1989)

However, the assumption that explicit negative feedback is needed for children to correct errors is entirely inconsistent with the principles of error-driven learning described above, and Ramscar and Yarlett (2007) provide an account of the way that general error-driven learning principles can give rise to the patterns of children's plural inflection acquisition. Ramscar and Yarlett's (2007) model represents plural items as semantic cues to phonological outcomes. Each item is an exemplar comprising an associatively linked semantic and a phonological component. For example, the plural noun CARS is represented by a couplet encoding the association between the general semantics of cars, including their plurality, and the phonological form */carz/*. The model assumes that learning is driven both by what the child has heard, and what the child expects to hear based on prior experience.

Over-regularization – children saying *foots* instead of *feet*, for example – arises in the model out of an initial failure to discriminate the individual semantic cues to particular plural words. In early learning, this lack of discrimination results in interference when shared cues activate frequent (and thus strongly learned) regular forms during the production of infrequent (and thus weakly learned) irregular forms. Interference thus results from prediction error generated by shared semantic cues. Accordingly, the associative values of these shared cues get weakened as learning progresses, which results in irregular forms becoming better discriminated and a decline in interference. Because regular and irregular forms are learned at different rates (there are far more regular than irregular plurals) and require different degrees of discrimination (regular plurals are supported by other regulars, but interfere with irregulars) the model predicts that interference effects will worsen for a time in the earliest stages of learning (because of the speed with which regular forms are learned), before slowly resolving as irregular forms become better learned. The model thus predicted that older children could improve their production of correct irregulars by repeatedly generating plurals (indeed even if they produce over-regularizations), but that this might be less beneficial to younger children.

These predictions were supported by the outcomes of several empirical tests (Ramscar & Yarlett, 2007). In one study, children repeatedly named plurals (correctly and incorrectly) for several blocks of regular and irregular items. The older children converged on the correct irregular plurals (e.g., production of "child" decreased, while "children" increased), without corrective feedback, however under the same conditions, younger children's over-regularization worsened, consistent with 'U-shaped' learning. A similar pattern of data was obtained when a semantic memory task for pictures was interspersed between pre- and post- tests of plural production: older children who performed an old/new task on pictures of regular and irregular plural items over-regularized less on the post-test, while younger children over-regularized more.

**Can over-regularization be reduced by exposure to regular items alone?**

A strong, very counterintuitive prediction that arises out of the principles of error-driven learning was not tested in Ramscar & Yarlett's (2007) studies. This is that at an appropriate point in learning, children's tendency to over-regularize irregular plurals will be reduced if they given training on only regular plurals. The way that this surprising prediction arises can be explained as follows: because regular nouns in English are frequent (both in terms of the number of regular plural noun types, and the overall number of plural noun tokens that are regular), the majority of plural forms cued by "plurality" will be plural forms which resemble their singular forms, but which end in + /S/. Since over-regularization is a failure to discriminate the appropriate cues to individual items present, (i.e., generalization) – if children encounter the cues of to regular plurals (e.g., a group of dogs), poor discrimination will result in the prediction of irregulars. The resultant prediction error will lead to children learning to negatively associate regular cues with irregular forms, which will increase the discrimination of regulars and irregulars. This increased discrimination of irregular plurals will in turn lead to a reduction in over-regularization. Further, although prediction errors for irregular items are caused by the activation of the cues for regular items, the erroneous prediction of irregulars is a function of how well the irregular items have been learned. Early in development, when irregulars are weakly learned, exposure to regular plurals will generate little irregular prediction error as compared to later in development, when irregulars will be better learned.

## Simulation Experiment

To formally test these ideas, we implemented a simple model of how children might learn to discriminate plural forms over time (see also Ramscar & Yarlett, 2007). The model assumes that plural items are represented as semantic cues to phonological outcomes. In early learning, over-regularization arises because the semantic representations of irregular plural items are not sufficiently discriminated from those of regular plurals, i.e., children initially tend to pluralize in response to general plurality, rather than in response to specific plural items (Ramscar & Yarlett, 2007). In the simulation, this was represented in terms of two competing hypotheses, which were reinforced whenever an irregular plural item was presented. One hypothesis was item specific (e.g., plural mouse is the cue to mice), while the other was more general (i.e., e.g., plurality is the cue to mice). Simultaneously, we simulated the learning of regular plurals. Due to the fact that regular plurals occur more frequently, and because their singular and plural forms overlap, we assumed that they offer more support to the general plural semantic hypothesis than irregular plurals, which instead offer support to more item-specific hypotheses.

Learning about the couplets was simulated using the Rescorla-Wagner (1972) rule described above. In the simulation, the learning rate, $\beta_j$ , for the semantic hypotheses (cues) was set at a constant, and $\lambda_j$ was set at 100% for the semantic-phonological couplets, which included both regular and irregular plurals forms. To simulate the high type and token frequency of regular plurals, $V_{ij}$ for the regular plurals was learned with $\alpha_i$ set to a high value (i.e., in the Rescorla-Wagner model, $\alpha_i$ effectively serves as a separate learning rate for each cue$_i$) while $V_{ij}$ for the irregular plurals was learned with $\alpha_i$ set to a low value.[2] This allowed training to be simulated by alternately presenting the model with regular and irregular items in training, to simulate a child's exposure to regular and irregular plurals at different frequency levels.

To examine the effect of exposure to regular plurals alone at different stages in learning, the presentation of irregular plurals was withheld for 10 trials, the first of

[2] In the simulation: $\beta_j$=0.3 $\alpha_I$ regular=0.4; $\alpha_I$ irregular=0.15.

these coming early in the model's training, and the second later in training, after the response to regular plurals had asymptoted. Figure 2 shows the learning of the two irregular hypotheses (general and specific) and the general regular hypothesis.

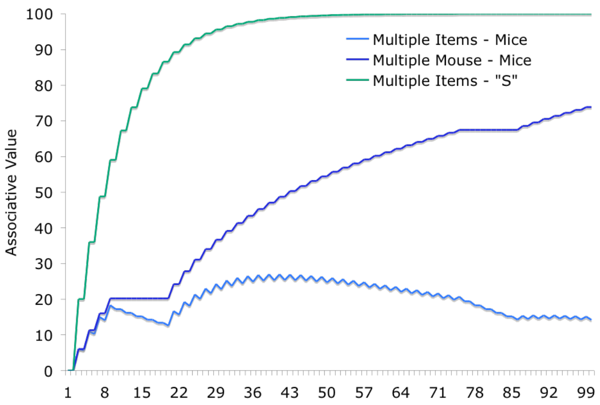


**Figure 2.** Learning of the semantic cues to an irregular item such as *mice* and the regular /S/. The periods in which no irregular trials occured appear as horizontal lines on the plot representing the *multiple mouse items⇒ mice* hypothesis.

As in Ramscar & Yarlett (2007) the likelihood of over-regularization (i.e. failure to produce the learned response) was modeled as a result of response competition, caused by spreading activation to items in memory that are activated by the semantics of the situation but which correspond to different phonological forms. This activation is modeled as a function of the degree to which the competing semantic-phonological couplets have been learned, the strength of the semantic cue that co-activates them and a spreading activation parameter *S* (Ramscar & Yarlett, 2007). Figure 3 shows the strength of this interference signal across the training period, and Figure 4 shows the effect this competition has on the likelihood that a learned irregular response will be reproduced. In Figure 4, response propensity is calculated by subtracting the value of the interference signal from the value of the correct response (Ramscar & Yarlett, 2007).

As can be seen from Figures 3 and 4, prediction errors for irregular items are caused by the activation of cues related to regular items, which results in the unlearning of the *multiple items⇒irregular* cue. Early in development, when irregulars are weakly learned, exposure to regular plurals will generate less overall irregular prediction error, and the overall frequency of regulars will result in a steady increase in the level of interference that produces over-regularization. Later in development, exposure to regular plurals produces more irregular prediction error, and interference no longer increases. As a result, the model predicts that depending on the overall prior exposure a child has had to plurals, exposure to regular plurals alone can lead to opposite effects (e.g., 'U-shaped' learning; Ramscar & Yarlett's 2007 model and empirical data showed that interspersing regular and irregular items produced this pattern of learning).

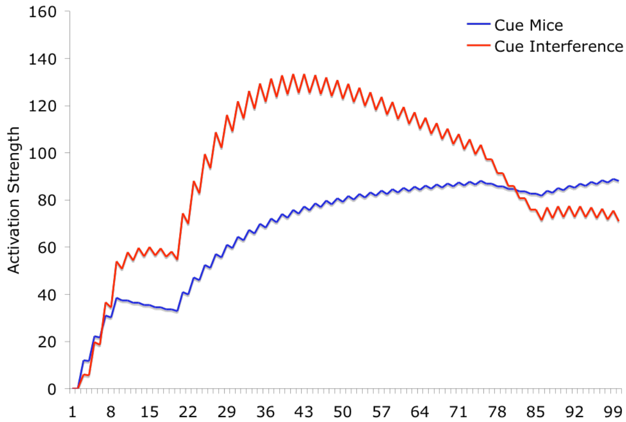


**Figure 3.** Interference and imitation in training. These parameter values were chosen to best illustrate our predictions; the important thing to note is the wunderlying *relationship* that arises out of the different learning rates.

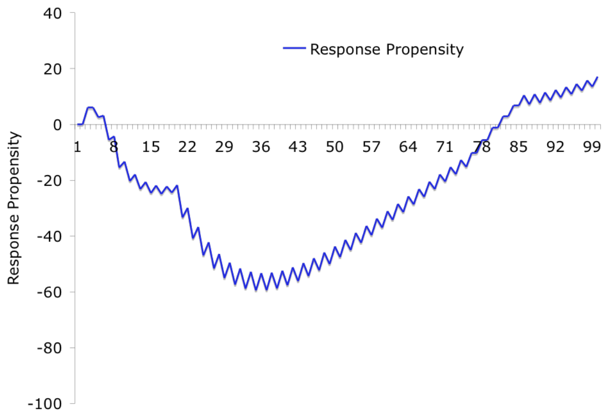


**Figure 4**. Response propensity levels over training. Over-regularization will be likely when this value is negative.

## Human Experiment

We tested these predictions using a semantic old/new task to expose children to regular plurals, and a test-train-test paradigm to establish a baseline rate of over-regularization for each child. This allowed us to examine the effect of children's exposure to regular plurals has on later irregular plural production (see Ramscar & Yarlett, 2007). Semantic priming (e.g., where priming the semantics of "doctor" yields shorter response latencies in a lexical decision task on "nurse";

Meyer & Schvaneveldt, 1971) indicates that phonological and orthographic representations can be activated by cueing their semantic features. The Ramscar & Yarlett (2007) model assumes that until the representation of a phonological–semantic association reaches asymptote, the activation of an association can strengthen its representation (see Roediger & Karpicke, 2006). Thus explicitly priming the semantics of the nouns, even in the absence of any overt naming responses by the child, was expected to be sufficient to produce errors in prediction and subsequent latent learning. Furthermore, by not having children explicitly name items, we aimed to reduced the effects of perseveration on spoken motor responses have in children's performance during a post-test. This we expected would allow for a better measure of their representation of the items tested.

**Participants**

24 four and 23 six year old children living resident in the vicinity of Palo Alto, California, and recruited from a database of volunteers. The average ages were 4 years and 7 months for the four year olds, and 6 years and 7 months for the six year olds.

**Methods and materials**

The children were randomly assigned to two groups, both of which were pre-tested on plural production.[3] In the elicitation test the children were asked to help a cookie monster puppet name a series of six irregular nouns, and six regular pairings of plural nouns. The children sat with the experimenter and named the nouns first from singular and then from plural depictions that were presented on a laptop computer.

In the experimental condition the children then performed an old/new task in which they were asked to tell a cookie monster whether or not they had seen depictions similar to those they had named in the pre-test. All depictions of the "old" items in training were novel, which required children to make categorization judgments to generate the correct answers. The children were asked to help the cookie monster identify them "By telling him, yes or no" to indicate whether they had already seen these depictions or not. When an object appeared, the experimenter asked the child to "Look at those – did cookie monster see those before?" Children who did not spontaneously respond were prompted, "Did cookie see these? Yes? No?". If no response was forthcoming, the experimenter proceeded to the next item. Half of the presented items were new depictions of the regular items in the pre-test and half were foils. The children were thus tested on 12 new and 12 old items per block. All of the items were presented as depictions on a computer screen.

In the control condition, the children were shown 6 color slides after the pre-test, and then asked to tell the cookie monster whether they had seen that particular color before in an old/new task that contained an equal number of foils. The colors were presented as blocks filling the computer screen to avoid cuing any notion of plurality. The total time to complete each was equal. Both sets of children were then post-tested on exactly the same set of depictions that were used in the pre-test.

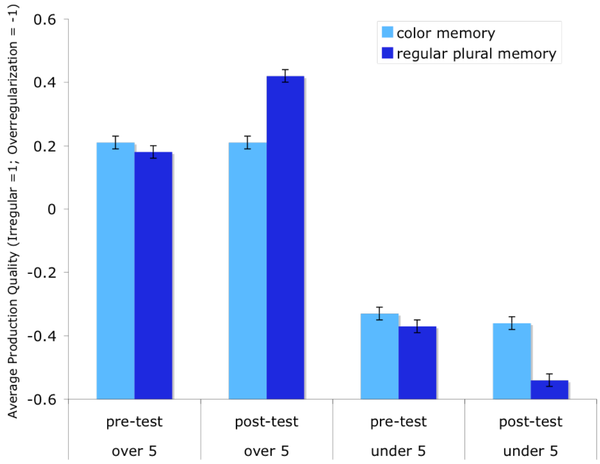


Figure 5. Pre and post test performance by age and condition

**Results**

The results overwhelmingly supported our predictions. The performance of the older children in the experimental condition improved between pre-and post test ($t(64)=2.256$, $p<0.05$) while the performance of the younger children declined ($t(66)=1.955$, $p<0.05$). There was little change in the performance of either age group in the control condition (see Figure 5). A 2 (pre- to post- test) x 2 (age) x 2 (condition) repeated measures ANOVA of the children's plural production revealed a significant interaction between age and pre- to post-test performance ($F(1,43) = 8.32$, $p<0.01$), and a significant interaction between age, training type and pre- to post-test performance ($F(1,266) = 4.235$, $p=.05$).

## General Discussion

We found that testing memory for regular plurals significantly reduced the rates of plural over-regularization in six-year-olds. Though the strength of these results is likely to have been influenced by recency (children named the irregulars immediately prior to regular training), what is clear that the children learned about irregular plurals, and improved their

[3] The irregular items were MOUSE-MICE, CHILD-CHILDREN, SNOWMAN-SNOWMEN, GOOSE-GOOSE, TOOTH-TEETH and FOOT-FEET; the regular matches were RAT, DOLL, COW, DUCK, EAR, and HAND. Ramscar & Yarlett (2007) Experiment 1 revealed that although children of these ages over-regularize these irregular plurals, they have reliable knowledge of their correct forms.

production of them, even though none were present during the training trials. We feel that, to the extent that this result is surprising, this surprise is due to the lack of widespread understanding of error-driven learning processes (see also Rescorla, 1988).

Overwhelmingly, research into language learning has pre-occupied itself with the observable: that is, with what a child hears or sees. Researchers have variously touted "the lack of negative evidence" in language learning as a constraint on theory (Marcus, 1993; Pinker, 2004), and much virtue is attributed to models that learn from "positive evidence" alone. We feel this is regrettable. There is good reason to believe that error-driven learning describes the principal mechanism by which people acquire information about their environment (Miller, Barnet & Grahame, 1995; Siegel & Allen, 1996; Ramscar & Yarlett, 2007; Ramscar, et al, in submission). The basic principles of error-driven learning are supported both by animal (e.g., Kamin, 1969; Rescorla & Wagner, 1972) and neurobiological models (e.g., Hollerman & Schultz, 1998; Barlow, 2001). In developing accounts of human learning, error-driven learning ought to be primarily considered when it comes to establishing conceptual and theoretical constraints and default hypotheses.

Extrapolating from the findings presented here (see also Ramscar & Yarlett, 2007; Ramscar et al, in submission), it seems likely that the processes involved in verbal learning – reducing prediction-error between semantic cues in the world and linguistic forms – are critical to the development of our use of language as an abstract representational device in communication.

Understanding language in terms of learning may, in the future, involve a reassessment of what human communication involves, requiring and inspiring new theories of language and its role in culture (Wittgenstein, 1953; Quine, 1960; Tomasello, 1999). At the very least, we would argue that simply reversing the trend ~~to~~ of ignoring learning in human development, we can and will reap many important scientific benefits.

## References


Baker, C. (1979). Syntactic theory and the projection problem. *Linguistic Inquiry*, 10, 533–581

Barlow H. (2001). Redundancy reduction revisited, *Network: Computation in Neural Systems*, 12, 241-253

Brown, R., & Hanlon, C. (1970). Derivational complexity and order of acquisition in child speech. In J. R. Hayes (Ed.), *Cognition and the development of language*. New York: Wiley.

Hollerman J.R., Schultz W. (1998) Dopamine neurons report an error in the temporal prediction of reward during learning. *Nature Neuroscience*, 1: 304-309.

Kamin L.J. (1969). Predictability, surprise, attention, and conditioning. In: Campbell B, Church R (eds). *Punishment and Aversive Behaviour*. Appleton-Century-Crofts: New York.

MacWhinney, B. (2004). A multiple process solution to the logical problem of language acquisition. *Journal of Child Language,* 31, 883–914

Marcus, G. F. (1993). Negative evidence in language acquisition. *Cognition*, 46, 53–85

Meyer, D.E. and Schvaneveldt, R.W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, 90, 227-234

Miller R.R., Barnet R.C. and Grahame N.J. (1995). Assessment of the Rescorla-Wagner Model, *Psychological Bulletin*, 117(3), 363-386.

Pinker, S. (1984). *Language learnability and language development.* Cambridge, MA: Harvard University Press

Pinker, S. (2004) Clarifying the logical problem of language acquisition. *Journal of Child Language*, 31, 949-953.

Quine, W.V. (1960) *Word and Object*, Cambridge, MA: MIT Press.

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

Ramscar, M. Yarlett, D., Dye, M., Denny, K., & Thorpe, K. (in submission) The Feature-Label-Order Effect In Symbolic Learning

Rescorla R.A. (1988). Pavlovian Conditioning: It's Not What You Think It Is, *American Psychologist*, 43(3), 151-160.

Rescorla R.A. and Wagner A.R. (1972). *A Theory of Pavlovian Conditioning: Variations in the Effectiveness of Reinforcement and Nonreinforcement.* In Black & Prokasy (Eds.), Classical Conditioning II: Current Research and Theory. New York: Appleton-Century-Crofts.

Roediger, H.L., & Karpicke, J.D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17, 249-255

Rudy, J. W. (1974). Stimulus selection in animal conditioning and paired-associate learning: Variations in the associative process. *Journal of Verbal Learning & Verbal Behavior,* 13, 282-296

Rumelhart D. E. & McClelland J. L (1986) On learning past tenses of English verbs. In Rumelhart & McClelland (eds) *Parallel Distributed Processing: Vol 2*: Psychological & Biological Models. Cambridge, MA: MIT Press.

Siegel, S.G., & and Allan, L.G. (1996). The widespread influence of the Rescorla-Wagner model, *Psychonomic Bulletin and Review*, 3(3), 314-321

Tomasello, M. (1999). The cultural origins of human cognition. Cambridge, MA: Harvard University Press

Wittgenstein, L (1953). *Philosophical Investigations* Oxford: Blackwell