Research Report

Children Value Informativity Over Logic in Word Learning

Michael Ramscar[1], Melody Dye[2], and Joseph Klein[3]
[1]Department of Linguistics, University of Tübingen; [2]Cognitive Science Program, Indiana University; and [3]Department of Psychology, Stanford University

Abstract
The question of how children learn the meanings of words has long puzzled philosophers and psychologists. As Quine famously pointed out, simply hearing a word in context reveals next to nothing about its meaning. How then do children learn to understand and use words correctly? Here, we show how learning theory can offer an elegant solution to this seemingly intractable puzzle in language acquisition. From it, we derived formal predictions about word learning in situations of Quinean ambiguity, and subsequently tested our predictions on toddlers, undergraduates, and developmental psychologists. The toddlers' performance was consistent both with our predictions and with the workings of implicit mechanisms that can facilitate the learning of meaningful lexical systems. Adults adopted a markedly different and likely suboptimal strategy. These results suggest one explanation for why early word learning can appear baffling: Adult intuitions may be a poor source of insight into how children learn.

Keywords
language, learning, language development, cognitive development, prediction

Received 2/27/12; Revision accepted 8/14/12

How do children figure out the meanings of the words they hear? How might a child learn that homes are *homes* and doors are *doors*, and not vice versa? The answer cannot be that children are more likely to hear the word *door* when doors are present, because people opening doors are more likely to say, "Hi, honey, I'm home!" than "I am now opening this door." Thus, it seems unlikely that a child could ever learn the meaning of a word simply by attending to how often that word is heard in tandem with an object or event. Indeed, hearing a word in the presence of an object tells a learner relatively little about its meaning: Though *door* could be the name of the object, it might equally likely relate to its color or texture, an action that could be taken upon it, or even a characteristic of the person knocking on it (Gleitman, 1990).

Here, we examine a possible solution to this problem proposed by Quine (1953), who suggested that rather than learning word meanings individually, children might instead discover how sensory experiences connect with systems of words. Concordantly, we found that in a novel word-learning task, children judged what was most informative about words (Shannon, 1948) by attending to how reliably the words co-occurred with objects and events in the environment relative to other competing matches. Why then have researchers traditionally focused on how children learn meanings in isolation (Carey & Bartlett, 1978; Heibeck & Markman, 1987)? It may be because that is what adults do: Faced with the same word-learning task, adults in our study adopted a logical strategy that treats meanings as determinate, individual entities. Gaining a better understanding of the way children learn word meanings, and the way their approach differs from that of adults, may offer important insights into what actually constitutes a rational approach to word learning.

Lessons From Learning Theory

Quine's proposal that children learn word meanings against the background of a system can be tested within the framework of modern learning theory. Experimental

Corresponding Author:
Michael Ramscar, Department of Linguistics, Eberhard Karls University, Tübingen, Germany
E-mail: michael.ramscar@uni-tuebingen.de

work in animal learning indicates that when animals are learning predictive relationships in their environment, they do not simply chart how often cues predict certain outcomes (reinforcement); they also track how often cues fail to predict potential outcomes (prediction error). The predictive value of any given cue is assessed against an entire network of cue-outcome relationships (Gallistel, 2003; Rescorla & Wagner, 1972), such that learning about that cue is ultimately driven more by implicit negative evidence (failures of alignment) than by positive evidence (Ramscar, Yarlett, Dye, Denny, & Thorpe, 2010; see also Fitneva & Christiansen, 2011).

For example, if rats are subjected to conditioning in which tones are followed by mild shocks, the rats will learn to respond fearfully to the tones. However, if tones that do not lead to expected shocks are added to the tone-shock pairings, the rats' conditioned responses will weaken in direct proportion to the increased background rate of tones (Rescorla, 1968). This weakening occurs because the rats' responses depend on how informative the tones are about the shocks (Kamin, 1969). Similarly, if children are sensitive to the value of information in word learning, then rather than simply tracking how often words and objects are paired together (e.g., a door is seen and *door* is heard), they might also track how often a potential pairing does not occur (e.g., a door is seen and *door* is not heard). By attending to the reliability of potential pairings, rather than simply tracking positive co-occurrences between words and objects, young word learners can effectively home in on which objects, actions, and events in their environment are most informative about which words.

Although it may seem surprising to suggest that a task as complex as word learning could be facilitated by such "dumb" mechanisms, error-driven learning results in far more sophisticated patterns of responses than is usually assumed (Rescorla, 1988), and a wealth of evidence points to the presence in humans of neural circuits that learn from error (Montague, Hyman, & Cohen, 2004; Waelti, Dickinson, & Schultz, 2001). It is also clear that executive function in young children differs markedly from executive function in adults, and that children's learning is less strategic and more information sensitive as a result (Ramscar & Gitcho, 2007; Thompson-Schill, Ramscar, & Chrysikou, 2009; Zelazo, Carlson, & Kesek, 2008). Thus, although simple learning models cannot accurately capture all the complexities of adult learning, they can offer particular insight into young children's learning. In recent years, these models have been fruitfully applied to the study of both linguistic development (Baayen, Milin, Filipovic Durdevic, Hendrix, & Marelli, 2011; Ellis & Sagarra, 2010; Ramscar, Dye, Popick, & O'Donnell-McCarthy, 2011; Ramscar, Dye, Witten, & Klein, 2013; Ramscar et al., 2010) and cognitive development (Colunga & Smith, 2005; Triesch, Teuscher, Deak, & Carlson, 2006).

Assuming that early lexical learning is an error-driven process, it should proceed smoothly as long as the words in a language are used in systematically informative ways. For example, provided that doors systematically have a higher co-occurrence rate (positive evidence) and lower background rate (negative evidence) for the word *doors* than for other less-reliable possibilities (e.g., homes, honeys, mailmen), then doors will be learned as the best predictor of the word *doors*. Careful studies of caregiver-child interaction indicate that child-directed speech is informative in precisely this way: The incidence of attentional matches—in which a child's attention is directed at an object when it is named—consistently exceeds that of mismatches, and there is little coherent covariation in the mismatches that do occur (see, e.g., Harris, Jones, & Grant, 1983).

A potential solution to the puzzle of how children learn words is that they learn in accordance with error-driven principles. Given that adult learning is complicated by conscious reasoning strategies, this solution would help explain why Quine's proposal is at odds with many of the standard approaches to language acquisition and with common adult intuitions about word meanings.

The Current Study

To test the merits of this proposal and examine the different ways in which informativity might inform word learning, we had children and adults participate in an ambiguous word-learning task with novel objects and labels, and with varying background rates of the objects. (For ease of presentation, in this overview we use one of the three different label sets used in the experiment.) Participants first saw two different novel objects, A and B, together and heard them labeled ambiguously as a *dax*. Subsequently, B was presented together with a new object, C, and another ambiguous label, *pid*. This training was repeated, and participants were then presented with all three objects and asked to identify the *dax*, the *pid*, or the *wug*, (a novel label not heard during training).

Because B occurred with both *dax* and *pid*, it had a higher background rate than either A or C; this made A most informative about *dax* and C most informative about *pid*. B's higher background rate also made it less informative about *wug* than either A or C, which were equally informative. A learner who was sensitive to the informativity of these pairings would therefore pick A as the *dax*, C as the *pid*, and either A or C as the *wug* (Fig. 1a). Conversely, a rational learner, employing a logical exclusion-based approach, might pick B as the *wug* (Fig. 1b).

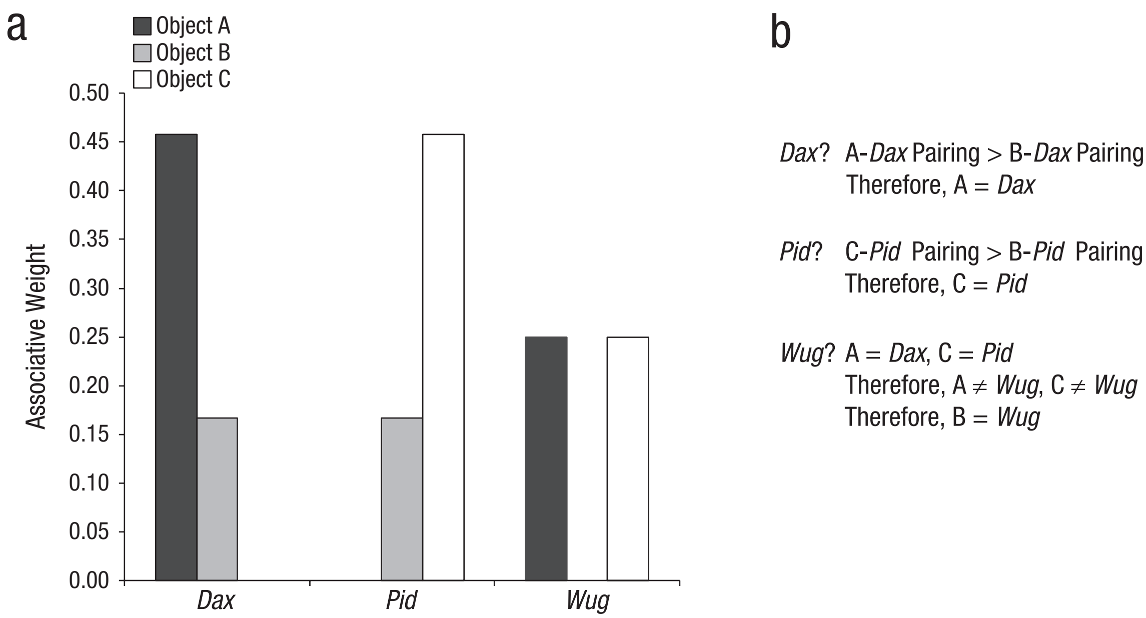


Fig. 1. Model predictions for the word-learning task in the experiment: (a) results of a formal simulation of the influence of background rates on learning and (b) predictions of a rational exclusionary approach to the same word-learning problem. The simulation in (a) used the equilibrium equations of the Rescorla-Wagner model (Danks, 2003; for analyses, see the Supplemental Material available online). The model has no free parameters and was given the same number of training trials as the participants in the present study. The model learned to associate object A most strongly with *dax*, and object C most strongly with *pid*, but associated both objects A and C with *wug*. In contrast, according to a rational exclusionary approach to this word-learning problem (b), given that A is the most likely *dax* and C the most likely *pid*, B must be the *wug*.

Method

Participants

Participants included 21 English-speaking children (12 girls, 9 boys) between 2 and 3 years of age (mean age = 28 months). In addition, 14 Stanford undergraduates and 20 developmental psychologists completed a version of the same task. The developmental psychologists surveyed were faculty and advanced doctoral students at leading research universities and were specialists in the study of children's language learning.

Materials

Three sets of objects, with three toys per set, were created from craft materials. The objects were designed to look like possible toys, without appearing too much like common objects. Within each set, the objects varied in size, color, and texture, to allow for easy discrimination. Pilot testing indicated that within each set, no particular object was consistently preferred over the other two objects. A set of syllable-matched novel words was paired with each set of objects, and matches were counterbalanced across subjects.

Procedure

Child version. The experiment consisted of a familiarization process, training, a short distraction period, and a recall test (Merriman, 1986). Training was administered using an interactive video format. This allowed for consistency of length and presentation, and controlled for unintentional social cues or attentional biases. Training, testing, and coding were conducted by hypothesis-blind experimenters.

Children were pretrained on the task using familiar objects. This familiarization period assessed participants' ability to choose among objects after first seeing them in a video. All participants "passed" this phase of training.

At the start of the training session for each set, the video's narrator (a puppet) announced that she would be showing the child some of her toys. First, objects A and B appeared on the video screen while the narrator used Label 1; then, objects B and C appeared while the narrator used Label 2. In both cases, the narrator used the labels conversationally, saying things like, "Do you see the dax? I really like the dax." In total, the narrator said each label nine times while the objects were visible. Finally, the narrator asked the child to repeat the label; the researcher paused the video at this point to allow the child to respond. If the child did not immediately respond, the researcher asked once more, and then resumed the video.

At the end of the training session for each set, the researcher removed the laptop computer used to play the video, and the child was asked if he or she would like to play another game. The researcher then retrieved a box containing all three objects the child had seen in the video. These interactions served as the distractor period. The researcher then asked the child to "show me the [target label]," repeating the question if the child was hesitant. The child was asked to respond to only one label—and hence, select one object—in each session. This was done for all three sets of objects, such that in training the child saw nine objects and heard six labels.

There were three test conditions: asking for Label 1; asking for Label 2; and asking for a novel label not heard in training, Label 3. Each child participated in all three conditions, with one condition per object set. The order of the conditions was counterbalanced across participants, and all participants were tested on each type of label only once. To conclude the experiment, the researcher repeated the three tests again, providing a measure of response consistency.

Undergraduate version. Undergraduate participants underwent the same training and testing as our 2- to 3-year-old participants did. They were tested individually and told that they were assisting in a pilot test of a task that was subsequently to be conducted with children. They were told that although the task might seem trivial, their answers were important and they should give the answers that seemed most natural to them.

Developmental-psychologist version. The design of the study was described in detail to each developmental-psychologist participant, who was then asked to predict how a healthy 2- or 3-year-old would respond. This survey was administered to assess expert opinion about how children would respond to this ambiguous word-learning task.

Results

Throughout the Results section, for ease of presentation, we use one set of labels—*dax, pid*, and *wug*—to refer to Labels 1, 2, and 3, respectively. From a purely informational perspective, A was a *dax*, C was a *pid*, and the same objects—A and C—were equally likely to be a *wug*. The 21 children we tested agreed: Their pattern of matching objects to labels matched well with the informativity of each object (Fig. 2). An analysis of variance (including data from the repeated tests) revealed a significant interaction of question (Label 1, 2, or 3) and object (A, B, or C), $F(1, 12) = 2.136$, $p < .025$. Object A was selected as the *dax* (M = 67%) with above-chance probability, $t(41) = 4.532$, $p < .001$; object C was selected as the *pid* (M = 62%) with above-chance probability, $t(41) = 3.421$, $p < .001$; and object B, which had the highest background rate, was selected as the *wug* (M = 17%) with below-chance probability, $t(41) = 2.858$, $p < .01$.

Although the children we tested matched objects to labels on the basis of informativity, the 14 Stanford undergraduates tested in exactly the same way did not. They agreed with the children about A and C, selecting A as the *dax* (M = 86%) and C as the *pid* (M = 79%) at above-chance levels, $t(13) = 5.401$, $p < .001$, and $t(13) = 3.421$,

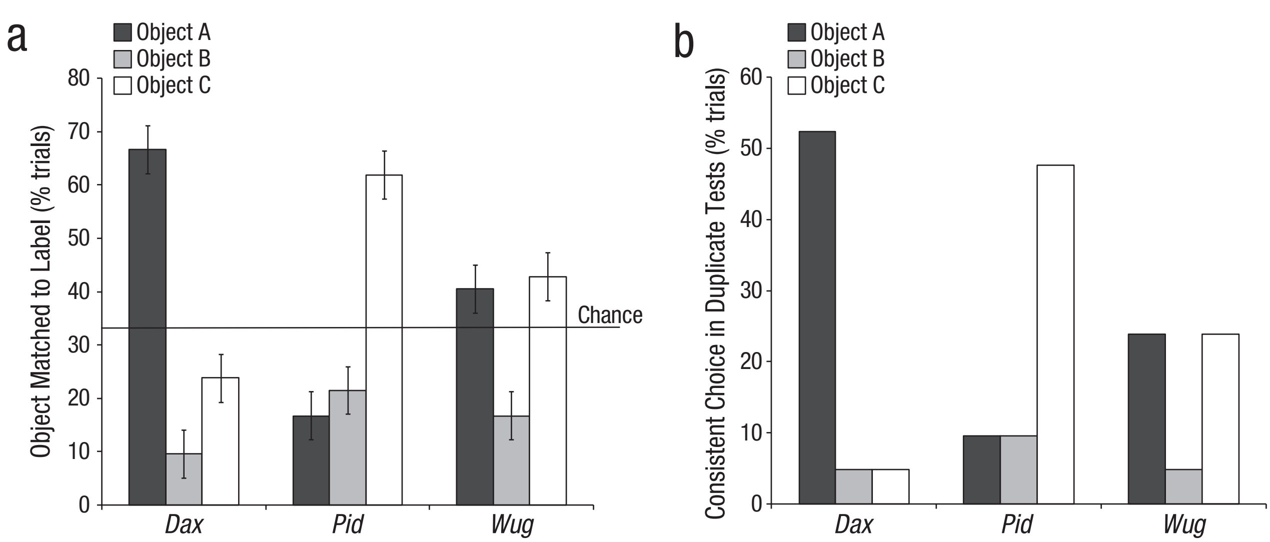


Fig. 2. Average percentage of trials on which the children (n = 21) selected each of the three objects as matching each label over repeated test trials (a) and the rate of consistent responses across the duplicate tests (b). Error bars represent standard errors of the mean.

$p < .01$, respectively. However, they chose B as the *wug* ($M = 64\%$) at above-chance levels, $t(13) = 2.332$, $p < .05$ (Fig. 3a). Further, although the group of developmental psychologists surveyed thought that the children would select A as the *dax* (M = 85%) and C as the *pid* (M = 95%), $t(19) = 6.311$, $p < .001$, and $t(19) = 12.34$, $p < .001$, respectively, they thought that the children would select B as the *wug* ($M = 80\%$), $t(19) = 5.089$, $p < .001$ (Fig. 3b). Thus, the psychologists predicted the undergraduates' behavior but not the children's behavior.

Discussion

The pattern of the children's responses indicates that children can and do use informativity when learning words. It appears that, as Quine suggested, the words children learn "face the tribunal of sense experience not individually but . . . as a corporate body" (p. 77). Children's word learning appears to be a systematic, rather than isolated, process, in which what is learned about any given word is dependent on its informativity in relation to other words and to context.

This pattern is consistent with recent findings in cross-situational studies of word learning, which have shown that children and adults can learn the meaning of words by "accruing statistical evidence across multiple and individually ambiguous word-scene pairings" (Smith & Yu, 2008, p. 1559). However, these findings, and many other similar findings in the lexical-acquisition literature, may provide only limited insight into the actual mechanisms underlying word learning because of the significant overlap in the predictions made by a number of qualitatively different theories (Yu & Smith, 2012). For example, in the classic mutual-exclusivity paradigm, 2-year-old children robustly match novel labels to novel objects rather than known objects (Merriman & Bowman, 1989; Merriman & Schuster, 1991; see also Mervis & Bertrand, 1994). Although these results are often taken as evidence that children are innately biased to assume that objects can have only one label, such results are equally consistent with learning from background rates.

In the present study, the well-specified nature of learning theory allowed us to derive predictions that discriminated between these alternatives. When we manipulated the background rates of several novel objects, we found no evidence of a bias toward mutual exclusivity—or other logical forms of inference—in 2- to 3-year-old children, who instead matched objects to labels depending on the objects' informativity in context: The same object that might be a *dax* in the context of a *dax* question was often matched to *wug* in the context of a *wug* question.

It is important to note that although associative learning is often mischaracterized in the wider psychological literature as co-occurrence counting, even animal learning is sensitive to prediction error and background rates (Rescorla, 1988). Allowing for similar learning mechanisms in children can make word learning appear far less baffling. For example, why do children map novel labels

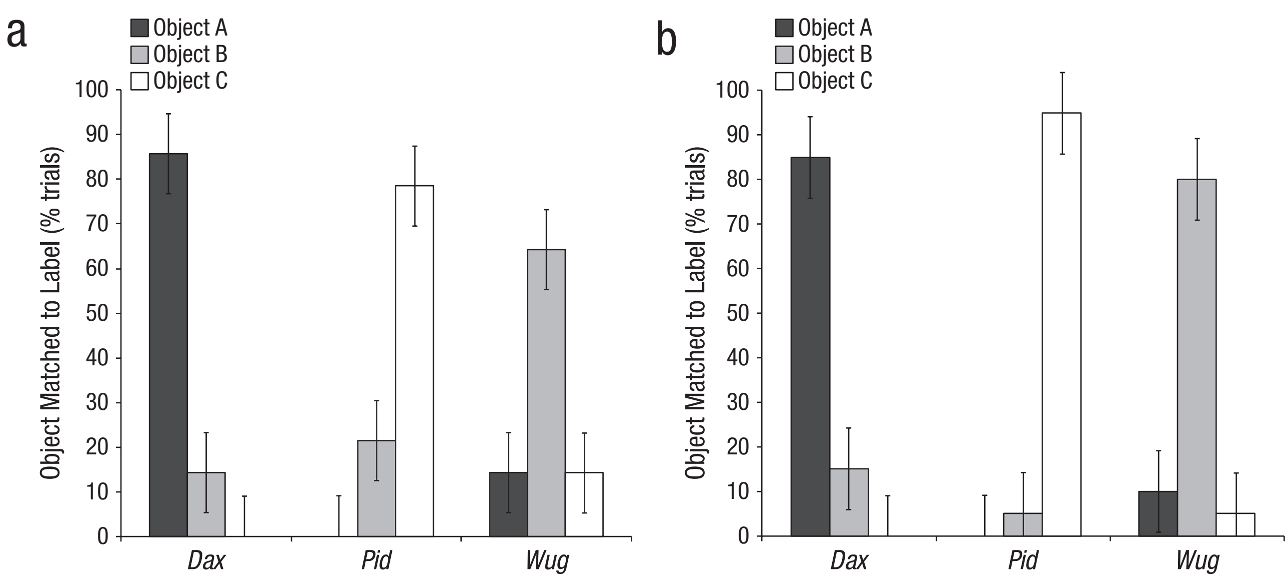


Fig. 3. Results from 14 Stanford undergraduates and 20 developmental psychologists. The graph in (a) shows the average percentage of trials on which the undergraduates selected each of the three objects as matching each label. The graph in (b) shows the psychologists' predictions for the percentage of trials on which they expected a healthy 2- or 3-year-old would select each of the three objects as matching each label. Error bars represent standard errors of the mean.

to objects (e.g., doors) and not to their features (e.g., brown and rectangular; Markman & Wachtel, 1988)? This question seems puzzling only if word learning is considered in isolation: From a learning-theoretic perspective, the answer lies in the simple fact that there are more brown and rectangular things in the environment than there are doors, so that doors (the objects) are more informative about the word *door* than are any of their properties. Indeed, because common properties like color have much higher background rates than typical objects do, color-word meanings are comparatively much harder to learn than object names (see Ramscar et al., 2010, for an account of early color learning). In a similar vein, Triesch et al. (2006) have shown how these same learning principles can explain how children learn to value the information in eye gaze, which helps them access the wealth of information about word meanings in the social environment.

Although the children in our study used informativity to match objects to labels, the adults did not. Instead, the adults appeared to employ a more strategic, exclusionary approach, reasoning that if objects A and C matched the labels seen in training, then the novel test label must apply to B. This may be a reasonable strategy in the restricted domain of our experiment, but in the real world, strict adherence to the logic of exclusion would fatally hinder learning multiple labels for an item. Given that adults do use labels flexibly—labeling the same person Margaret Thatcher, the prime minister, a wife, a mother, a member of parliament, or the Iron Lady, depending on the context—it seems likely that the performance of our adult participants does not reflect the way they learned the meanings of the words they use in everyday life (although it may help explain why adults struggle with second-language learning; see Arnon & Ramscar, 2012).

These results reveal not only significant differences in the way children and adults approached the same task, but also a tendency on the part of experts to assume that the approaches of children and adults would be the same. The first finding should come as little surprise; it is clear that adults process information in qualitatively different ways than young children (Hudson Kam & Newport, 2005; Ramscar & Gitcho, 2007; Siegler, 1994). What may be more surprising is how often adults fail to take this distinction into account.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

References

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

Baayen, R. H., Milin, P., Filipovic Durdevic, D., Hendrix, P., & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review*, *118*, 438–481.

Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Papers and Reports on Child Language Development*, *15*, 17–29.

Colunga, E., & Smith, L. B. (2005). From the lexicon to expectations about kinds: A role for associative learning. *Psychological Review*, *112*, 347–382.

Danks, D. (2003). Equilibria of the Rescorla-Wagner model. *Journal of Mathematical Psychology*, *47*, 109–121.

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

Fitneva, S. A., & Christiansen, M. H. (2011). Looking in the wrong direction correlates with more accurate word learning. *Cognitive Science*, *35*, 367–380.

Gallistel, C. R. (2003). Conditioning from an information processing perspective. *Behavioural Processes*, *62*, 89–101.

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

Harris, M., Jones, D., & Grant, J. (1983). Nonverbal context of mothers' speech. *First Language*, *4*, 21–30.

Heibeck, T. H., & Markman, E. M. (1987). Word learning in children: An examination of fast mapping. *Child Development*, *58*, 1021–1034.

Hudson Kam, C. L., & Newport, E. L. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development*, *1*, 151–195.

Kamin, L. J. (1969). Predictability, surprise, attention, and conditioning. In B. Campbell & R. Church (Eds.), *Punishment and aversive behavior* (pp. 279–296). New York, NY: Appleton-Century-Crofts.

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

Merriman, W. E. (1986). Some reasons for the occurrence and eventual correction of children's naming errors. *Child Development*, *57*, 942–952.

Merriman, W. E., & Bowman, L. L. (1989). The mutual exclusivity bias in children's word learning. *Monographs of the Society for Research in Child Development*, *54*(3–4, Serial No. 220), 1–132.

Merriman, W. E., & Schuster, J. M. (1991). Young children's disambiguation of object name reference. *Child Development*, *62*, 1288–1301.

Mervis, C. B., & Bertrand, J. (1994). Acquisition of the novel name–nameless category (N3C) principle. *Child Development*, *65*, 1646–1662.

Montague, P. R., Hyman, S. E., & Cohen, J. D. (2004). Computational roles for dopamine in behavioural control. *Nature*, *431*, 760–767.

Quine, W. V. O. (1953). *From a logical point of view*. Cambridge, MA: Harvard University Press.

Ramscar, M., Dye, M., Popick, H. M., & O'Donnell-McCarthy, F. (2011). The enigma of number: Why children find the meanings of even small number words hard to learn and how we can help them do better. *PLoS ONE*, *6*, e22501. Retrieved from http://www.plosone.org/article/info:doi/10.1371/journal.pone.0022501

Ramscar, M., Dye, M., Witten, J., & Klein, J. (2013). Dual routes to cognitive flexibility: Learning and response conflict resolution in the dimensional change card sort task. *Child Development*. Advance online publication. doi:10.1111/cdev.12044

Ramscar, M., & Gitcho, N. (2007). Developmental change and the nature of learning in childhood. *Trends in Cognitive Sciences*, *11*, 274–279.

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

Rescorla, R. A. (1968). Probability of shock in the presence and absence of CS in fear conditioning. *Journal of Comparative and Physiological Psychology*, *66*, 1–5.

Rescorla, R. A. (1988). Pavlovian conditioning: It's not what you think it is. *American Psychologist*, *43*, 151–160.

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. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York, NY: Appleton-Century-Crofts.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell Systems Technical Journal*, *27*, 379–423, 623–665.

Siegler, R. S. (1994). Cognitive variability: A key to understanding cognitive development. *Current Directions in Psychological Science*, *3*, 1–5.

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*, 1558–1568.

Thompson-Schill, S., Ramscar, M., & Chrysikou, E. (2009). Cognition without control: When a little frontal lobe goes a long way. *Current Directions in Psychological Science*, *8*, 259–263.

Triesch, J., Teuscher, C., Deak, G. O., & Carlson, E. (2006). Gaze following: Why (not) learn it? *Developmental Science*, *9*, 125–147.

Waelti, P., Dickinson, A., & Schultz, W. (2001). Dopamine responses comply with basic assumptions of formal learning theory. *Nature*, *412*, 43–48.

Yu, C., & Smith, L. B. (2012). Modeling cross-situational word-referent learning: Prior questions. *Psychological Review*, *119*, 21–39.

Zelazo, P. D., Carlson, S. M., & Kesek, A. (2008). The development of executive function in childhood. In C. Nelson & M. Luciana (Eds.), *Handbook of developmental cognitive neuroscience* (2nd ed., pp. 553–574). Cambridge, MA: MIT Press.

Supplementary Materials

Modeling learning

Learning in the training study simulated using the equilibrium equations (Danks, 2003) for the Rescorla & Wagner (1972) model. In the Rescorla - Wagner model, learning changes the associative strengths between cues C_i and an outcome O_j as the result of discrete learning trials. If the presence of a cue or outcome *X* at time *t* is defined as present(*X, t*), and its absence as absent(*X, t*), then the predictive value *V* of a cue *i* for an outcome *O* after a learning event at time *t + 1* can be stated as:

$$V_i^{t+1} = V_i^t + \Delta V_i^t$$

while the change (Δ) in the predictive value of *i* after *t* can be defined as:

$$\Delta V_i^t = \begin{cases} 0 & \text{if } ABSENT(C_i,\ t) \\ \alpha_i \beta_1 (\lambda - \sum_{PRESENT(C_j,t)} V_j) & \text{if } PRESENT(Cj,t) \,\&\, PRESENT(O,t) \\ \alpha_i \beta_2 (0 - \sum_{PRESENT(C_j,t)} V_j) & \text{if } PRESENT(Cj,t) \,\&\, ABSENT(O,t) \end{cases}$$

Thus, learning is governed by a *discrepancy function* where λ is the total value of the predicted event (i.e., the maximum amount of associative strength that an outcome *j* can support; here it is simply set to 1, indicating that an event is fully anticipated) and V_j is the predictive value for outcome *j given* the set of cues present at time *t*.

Although the trial-based implementation of the Rescorla-Wagner model does not capture the effects of pre-exposure of cues (see Miller, Barnett & Grahame, 1995), Danks (2003) presents a set of equations that allow the values learned across trials to be estimated by examining the system when it is in a state of equilibrium (i.e., once learning has finished), and that are sensitive to the effects that different levels of exposure – i.e., background rates – have on learning.

The association strengths Vi of the cues C to a specific outcome O can be obtained by solving the following system of equations, where n + 1 denotes the number of different cues (input features) and where the indices i and j range over the different cues:

$$\begin{pmatrix} \Pr(C_0 | C_0) & \Pr(C_1 | C_0) & \ldots & \Pr(C_n | C_0) \\ \Pr(C_0 | C_1) & \Pr(C_1 | C_1) & \ldots & \Pr(C_n | C_1) \\ \ldots & \ldots & \ldots & \ldots \\ \Pr(C_0 | C_n) & \Pr(C_1 | C_n) & \ldots & \Pr(C_n | C_n) \end{pmatrix} \begin{pmatrix} V_0 \\ V_1 \\ \ldots \\ V_n \end{pmatrix} = \begin{pmatrix} \Pr(O | C_0) \\ \Pr(O | C_1) \\ \ldots \\ \Pr(O | C_n) \end{pmatrix}$$

$\Pr(C_j|C_i)$ represents the conditional probability of cue C_j given cue C_i, and $\Pr(O| C_i)$ the conditional probability of outcome O given cue C_i. Informally, we can think of the association strengths V_j as optimizing the conditional outcomes given the conditional probabilities characterizing the input space. In this model, the estimation of the weights on the connections from cues to outcomes is parameter-free, and totally determined by the training data.

Simulations

To formalize the predicted effects of the training study, we simulated learning in response to the conditions examined in an open source program that implements the Danks equations: the Naïve Discriminative Learner Model (NDL; see Baayen et al. (2011) that runs a package in the “R” statistical computing environment, see:

cran.r-project.org/web/packages/ndl

The simulations exactly reflected the rates of exposure to the objects and labels in training and testing. In every training trial, each of the two trained labels was repeated nine times in the presence of two of the objects, with one object always being present, and the other two objects being present half the time. There was only one test trial per set of items: in these trials, all three objects were present, and one of the labels was given. Since there is strong evidence that learning occurs on test trials (Roediger, McDermott, & McDaniel, 2011), these were also included in the simulations.

Training and testing thus yielded 3 distinct conditions:

Condition 1

Objects Present	Label Heard	Label Frequency
Object A Object B	Label 1	9
Object B Object C	Label 2	9
Object A Object B Object C	Label 1	1

Condition 2

Objects Present	Label Heard	Label Frequency
Object A Object B	Label 1	9
Object B Object C	Label 2	9
Object A Object B Object C	Label 2	1

Condition 3

Objects Present	Label Heard	Label Frequency
Object A Object B	Label 1	9
Object B Object C	Label 2	9
Object A Object B Object C	Label 3	1

The objects were entered into the NDL model as cues (along with four cues to represent the experimental context, and the contexts of the different training and testing effects), while the labels were coded as outcomes (see Ramscar et al, 2010). These encodings are given in full in the appendix.

Results
The simulations produced the following patterns of learning:

Condition 1 (Trained: Labels 1 and 2; Tested: Label 1; Label 3 Not Presented)

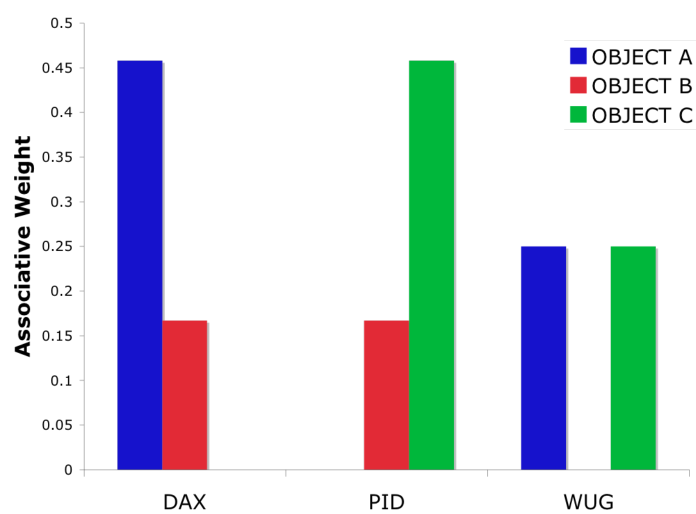
	Label 1	Label 2
ObjectA	0.458	-0.292
ObjectB	0.167	0.167
ObjectC	-0.042	0.208

Condition 2 (Trained: Labels 1 and 2; Tested: Label 2; Label 3 Not Presented)

	Label 1	Label 2
ObjectA	0.208	-0.042
ObjectB	0.167	0.167
ObjectC	-0.292	0.458

Condition 3 (Trained: Labels 1 and 2; Tested: Label 3)

	Label 1	Label 2	Label 3
ObjectA	0.208	-0.292	0.25
ObjectB	0.167	0.167	0
ObjectC	-0.292	0.208	0.25



Supplementary Figure1: Summary of simulation results.

As can be seen, as a result of Training Condition 1, Label 1 becomes most strongly associated with Object A, in Training Condition 2, Label 2 becomes most strongly associated with Object C, and in Training Condition 3, Label 3 is most strongly associated with Objects A and C.

References

Baayen, R. H., Milin, P., Filipovic Durdevic, D., Hendrix, P., & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review* 118, 438-482.

Danks, D. (2003). Equilibria of the Rescorla-Wagner model. *Journal of Mathematical Psychology*, 47(2), 109–121.

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

Ramscar, M., Yarlett, D., Dye, M., Denny, K., & Thorpe, K. (2010) Feature-Label-Order effects and their implications for symbolic learning. *Cognitive Science*, 34(7), 909-957

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., McDermott, K.B., & McDaniel, M.A. (2011). Using testing to improve learning and memory. In M.A. Gernsbacher, R. Pew, L. Hough, & J.R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 65-74). New York: Worth Publishing Co.

Appendix – NDR encodings for the three simulations

```
<Mido1.txt>
Cues                                          Outcomes    Frequency
ObjectA_ObjectB_Context1_ExptContext          Dax                 9
ObjectB_ObjectC_Context2_ExptContext          Pid                 9
ObjectA_ObjectB_ObjectC_Context3_ExptContext  Dax                 1

<Mido2.txt>
Cues                                          Outcomes    Frequency
ObjectA_ObjectB_Context1_ExptContext          Dax                 9
ObjectB_ObjectC_Context2_ExptContext          Pid                 9
ObjectA_ObjectB_ObjectC_Context3_ExptContext  Pid                 1

<Mido3.txt>
Cues                                          Outcomes    Frequency
ObjectA_ObjectB_Context1_ExptContext          Dax                 9
ObjectB_ObjectC_Context2_ExptContext          Pid                 9
ObjectA_ObjectB_ObjectC_Context3_ExptContext  Wug                 1
```

Ndr code

```
library(ndl)

dat<-read.table("midoX.txt",T,stringsAsFactors=FALSE)

###command for the equilibria of the association weights
w=estimateWeights(dat)
w=round(w,3)
```