Informativity versus logic: Children and adults take different approaches to word learning

Michael Ramscar, Melody Dye, Joseph Klein, Nicole Aguirre, Linda Ruiz & Lily Sadaat
Department of Psychology, Stanford University,
Jordan Hall, Stanford, CA 94305

Abstract

The question of how children learn what words mean is one that has long perplexed philosophers and psychologists. As Quine famously pointed out, the problem of accounting for word learning is a deep one: simply hearing a word uttered in the presence of an object tells a learner next to nothing about its meaning. Yet somehow, children learn to understand and use words correctly. How? Here, we find that learning theory offers an elegant solution to this seemingly intractable puzzle in language acquisition. To test its predictions, we administered an ambiguous word-learning task to toddlers, undergraduates and developmental psychologists. Intriguingly, while the toddlers' performance was consistent with our hypothesis – and with the workings of general learning mechanisms that would facilitate verbal acquisition – adult performance differed markedly. These results have implications both for how our adult intutions inform the study of early language learning and for problems in second-language acquisition.

Keywords: Word Learning, Error-Driven Learning, Learning Theory, Discrimination Models, Language Acquisition

Introduction

How do children figure out the meanings of the words they hear? How does a child learn that homes are "homes" and doors are "doors," and not vice versa? The answer cannot simply be that children are more likely to hear "door" when doors are present, because people opening doors are more likely to say, "Hi Honey, I'm home!" than "*I am now opening the door*." Given this, 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 (Gleitman, 1990). 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 relate to its color or texture, an action that could be taken upon it, or even a characteristic of the person knocking on it (Quine, 1953)

Here we examine a possible solution to this problem proposed by the philosopher W.V.O. Quine, who suggested that rather than learning word meanings individually, children must instead discover how sensory experience connects up with *systems* of words (see also Wittgenstein, 1953). In line with this suggestion, we find that in a novel word learning task, children judge what is most informative about words (Shannon & Weaver, 1949), by attending to the signal-to-noise ratio in their environment. Why then have researchers traditionally focused on how children learn 'meanings' in isolation? (see Smith & Yu, 2008 and Akhtar & Montague, 1999 for discussion) It may be because that's what adults do: faced with the same task, adults adopt 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, can help us better our approaches to teaching the young, while offering insight into the struggles many adults encounter with second language acquisition.

A Puzzle for Word Learning

The dilemma a child faces in word learning has often been framed as a classic induction problem. Faced with a novel word, the child must select from among multiple – perhaps even infinite – competing hypotheses as to what the word means, on the basis of relatively little evidence from the input (Carey, 1978; Bloom, 2000). This apparent philosophical conundrum has long been a source of puzzlement for child development researchers, because in spite of the presumed difficulty in narrowing the hypothesis space, children prove remarkably adept word learners.

This puzzle – of how children can learn so rapidly and so successfully despite the difficulties posed by 'referential uncertainty' – has led many researchers to posit native constraints on word learning. Proposals in this vein have ranged from innate concepts and conceptual primitives (Chomsky, 2000; Fodor, 1988), to syntactic bootstrapping (Brown, 1957; Landau & Gleitman, 1985; Naigles, 1990) and strong representational biases (Carey, 1978; Waxman & Gelman, 1986).

Though there is certainly much to distinguish these approaches, they share a common focus on high-level constraints, which are meant to meaningfully generalize across linguistic development and behavior. While such constraints may be useful in describing how children tend to behave as they are learning language, they do little to illuminate the underlying learning processes. Constraints still require an explanation involving either innate linguistic principles or another underlying mechanism that allows humans to learn (or otherwise deduce) these principles (Smith, 1995). Yet many theorists in this tradition have been satisfied to speculate that these default assumptions exist, without attempting to flesh out how they might be computationally or neurobiologically instantiated (for critiques, see Nelson, 1988; Rakison & Lupyan, 2008; Smith, Colunga & Yoshida, 2010; Ramscar et al., 2010).

To summarize, then, there has been considerable debate over both how word learning is conceptualized and understood, and whether proposed constraints are psychologically real constructs that restrict and delimit learning, or underspecified descriptive generalizations that may obscure underlying processes.

Quine's Proposal

While many theories of word learning seek to explain how children learn isolated words, Quine proposed that children learn the meanings of words against the background of a system, an idea that is consistent with the general frameworks of both learning and information theory. Experimental work in animal learning indicates that when learning the relationship between a cue and an outcome, animals 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 a cue is assessed against an entire system of cue-outcome relationships.

To give a simple example, if a rat is subjected to conditioning in which a series of tones is followed by mild shocks, the rat 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, rats' conditioned responses will weaken in direct proportion to the increased **background rate** of tones (Rescorla, 1968). This is because rats' responses depend on how *informative* the tones are about the shocks (Kamin, 1969; Rescorla & Wagner, 1972; Rescorla, 1988).[1]

Similarly, if children are sensitive to the value of information in word learning, than rather than simply tracking how often words and objects are paired together (e.g., a door is seen and "door" is heard), children might also track how often a potential pairing does not occur (e.g., a door is seen and "home" is not heard). By attending to the signal-noise ratio in the surrounding linguistic environment, they could home in on which objects, actions and events in the world are most informative about which words.

Error-Driven Learning

Why investigate how children learn words from the vantage point of animal learning? First, there is a wealth of evidence to support the idea that the neural mechanisms that underpin error-driven learning in animals are present in humans, and that they provide us with the same functional capabilities that are seen and predicted by animal models (Schultz, Dayan & Montague, 1997; Waelti, Dickinson & Schultz, 2001; Montague, Hyman & Cohen, 2004; Samejima et al., 2005; Colunga & Smith, 2005; Ramscar & Yarlett, 2007; Ramscar et al., 2010). Second, and perhaps more critically, prior research has made clear that adults' executive function differs markedly from that of children, and as a result, adult learning is typically far more strategic and less information-sensitive (Derks & Paclisanu, 1967; Ramscar & Gitcho, 2007; Thompson-Schill, Ramscar & Chrysikou, 2009). Thus, while simple error-driven models of learning might not accurately capture adult behavior in all instances, they could well provide key insights into how children learn words.

Assuming an error-driven process, word learning should proceed smoothly so long as the words of a language are systematically informative. For example, provided that doors have a higher co-occurrence rate with the word "doors" (positive evidence) and a lower background rate (negative evidence), compared to other less reliable possibilities (such as homes, Honeys or mailmen), then an error-driven model will learn that doors are most informative about "doors" (for a review, see Ramscar et al., 2010). If children co-opt error-driven learning mechanisms for the purposes of learning words, it would offer a potential solution to the word-learning puzzle posed at the outset. At the same time, assuming that adults don't do this, it would help explain why Quine's proposal is at odds with many of the standard approaches adult researchers have devised to study language acquisition (Carey, 1978; Markman, 1989; Bloom, 1994), and with common adult intuitions about the nature of word meanings.

Experiment

To test the merits of this proposal – and examine the different ways in which informativity might 'inform' word learning – we trained children and adults on novel word meanings while manipulating the background rates of the objects paired with the labels that they learned. Our participants first saw two different novel objects together (A and B) and heard them labeled ambiguously as a "DAX" (Figure 1). Subsequently, B was presented with a new object, C, and another ambiguous label, "PID." This training was repeated, and the participants were then presented with all three objects, and asked to identify either the "DAX," the "PID," or the "WUG," which they hadn't heard before.

Because B occurs with both "DAX" and "PID," it has a higher background rate than either A or C, which makes A more informative about "DAX," and C more informative about "PID." Critically, B's higher background rate also makes it less informative about the novel word "WUG" than A or C, which are both equally informative about "WUG." From a purely informational perspective, then, A is a DAX, C is a PID, and A or C are WUGs (Rescorla, 1968).

Here, we tested whether our participants were sensitive to information in learning, or whether they adopted a more 'logical' approach, and paired B with the novel word "WUG." (Which would be consistent with the proposed 'mutual exclusivity' constraint on word learning, which holds that objects that don't have names will be the most likely candidates for mapping to a new label; Markman, 1989).

To assess the nature of our subjects' approaches to word learning, each participant received training on 3 different sets of objects and words, and was tested at the end of each training session, and again at the end of the experiment.

[1] "By itself contiguity between a CS and US is insufficient for Pavlovian conditioning. Rather, for a CS to become conditioned, it must in some sense provide *information* about the coming of the US; the CS must not only be paired with the US, it must *predict* its coming." (Rescorla, 1972)

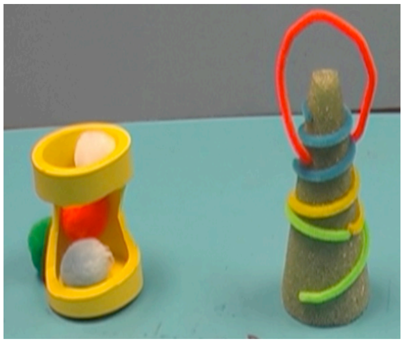


Figure 1. Sample objects used in Training. The objects were varied in shape, color, and texture to allow discrimination, and counterbalanced across our participants to control for attractiveness.

Participants

21 English-speaking children between 2- and 3- years (M age = 28 months old) participated in this study, with a near even balance between genders (12 girls, 9 boys). All children participants were recruited from Stanford and the surrounding community. In addition, there were two groups of adult participants: 14 Stanford undergraduates and 20 Developmental Psychologists. The Developmental Psychologists surveyed were faculty and advanced doctoral students at leading research universities specializing in the study of children's language learning.

Materials

3 sets of objects, with 3 toys per set, were created from craft materials. The objects were designed to look like possible toys, without appearing too much like any common objects. Within each set, the objects varied in size, color, and texture, allowing for easy discrimination between each object. Pilot testing indicated that within each set, no particular object was consistently preferred to the other objects.

A set of syllable-matched novel words was paired with each set of objects, and matches were counterbalanced across subjects.

Procedure

The experimental design was modeled on classic word learning studies in young children (Merriman, 1986; Woodward et al, 1994), and consisted of: familiarization, training, short distraction, and a recall test. Training, testing and coding was conducted by hypothesis-blind experimenters.

Notably, pilot testing indicated that when children were presented with physical objects, they would sometimes reach for the objects or attempt to play with one or more during the training session. To avoid biased attention towards any particular object during training, the training was conducted using a narrated video. Using video training also allowed for consistency of length and presentation, and controlled for unintentional social cues (such as eye gaze).

Familiarization

Children were pre-trained on the task using familiar objects. The first video clip presented two common household objects (a cup and a pair of sunglasses). While both objects were onscreen, the narrator talked about the cup, and then told the child that "my friend" (meaning the researcher) had some similar objects, and that they would now play with those objects. The researcher paused the video and placed the cup and sunglasses on the table in front of the child. The researcher then asked the child to show her the cup. Once the child made a choice, the child was allowed to play with both objects briefly. This familiarization period was designed to make the participants feel comfortable choosing between physical objects after first seeing them in the video.

All the participants tested answered the familiarization question correctly and readily, suggesting that the children understood the nature of the task, and that switching from video to real objects was not a barrier to performance.

Training

At the start of the training session for each set, the puppet welcomed the child and announced that she would be showing the child some of her toys. First, Objects A and B would appear on screen while the narrator used Label 1 (e.g., *DAX*); then, Objects B and C would appear while the narrator used Label 2 (e.g., *PID*). In both cases, the narrator would use the Labels conversationally, saying things like "Do you see the Dax? I really like the Dax." To keep the child engaged, the puppet also played a game with the toys on screen, hiding them and then bringing them back out for the child to see again. In total, the puppet said the Label nine times while the objects were visible. Additionally, the puppet asked the child to repeat the Label; the researcher paused the video at this point to allow the child to respond. If the child didn't immediately respond, the researcher asked once more, and then resumed the video.

At the end of each training session, the researcher stopped the video, moved the screen off of the table, and brought out all three objects. The researcher then asked the child to "**show me the [target label]**," and repeated the question again if the child was hesitant. Once the child chose an object, the researcher recorded it and encouraged the child to play with each of the objects briefly, before moving on to the next training session. This was done for 3 sets of objects (3 training and testing sessions), such that the child learned about 6 labels and 9 objects.

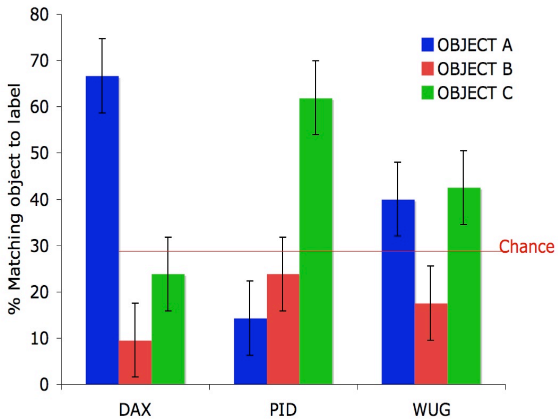
Conditions

There were three test conditions: asking for *Label 1* (e.g., DAX), asking for *Label 2* (e.g., PID), or asking for a novel label, not heard in training, *Label* 3 (e.g., WUG). Each child participated in all three conditions, with one condition per object set. The order of the conditions was counterbalanced across subjects, and all subjects were tested on each type of label only once. To conclude the experiment, the researcher repeated the three tests again, providing a second measure of learning.

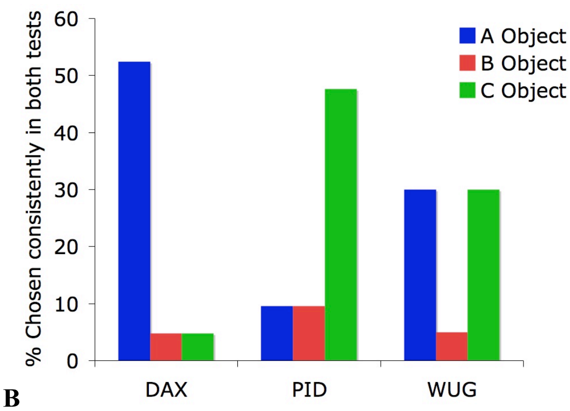
Results

From a purely informational perspective, A is a DAX, C is a PID, and A or C are WUGs. The 21 children (12 girls, 9 boys, M age =28 months) we tested agreed: their pattern of matching objects to labels matched exactly with the informativity of each object. ANOVA (Question x Object) = $F(1,12)=2.136$, $p<0.025$); P(DAX=A > chance, M=.67), $t(41)=4.532$, $p<0.001$; P(PID=C > chance, M=.62), $t(41)=3.421$, $p<0.001$; P(WUG=B < chance, M=.17), $t(41)=2.858$, $p<0.01$).

Notably, while the children we tested matched objects to labels on the basis of informativity (Figure 2A), 14 Stanford undergraduates we tested in exactly the same way did not. They agreed with the children about A and C (P(DAX=A > chance, M=.86) $t(13)=5.401$, $p<0.001$; P(PID=C > chance, M=.79), $t(13)=3.421$, $p<0.01$), but chose B as the WUG (P(WUG=B > chance, M=.64), $t(13)=2.332$, $p<0.05$; Figure 2B), which is the opposite of what the children did. Further, when we surveyed a group of Developmental Psychologists and asked them to predict children's behavior in our task, they too thought B was the WUG (P(WUG=B > chance, M=.80), $t(19)=5.089$, $p<0.001$; P(DAX=A > chance, M=.85) $t(19)=6.311$, $p<0.001$; P(PID=C > chance, M=.95), $t(19)=12.34$, $p<0.001$; Figure 2C): meaning they correctly predicted the behavior of the undergraduates, but not the children.

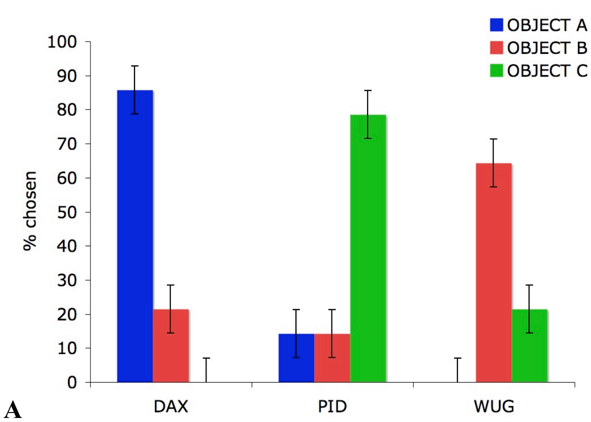


A

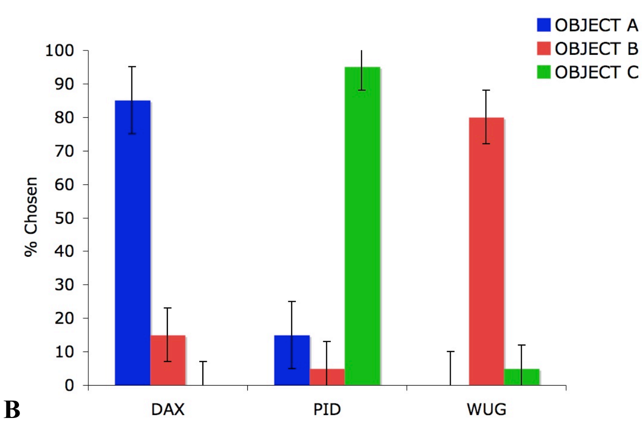


B

Figure 2. Responses from 21 children. Object B, which had the highest background rate, was chosen at below chance levels across all of the trails, including the critical "wug" trial. 2A shows average responses over all of the tests, while 2B shows the rate of consistent responses across the duplicate test trials.



A



B

Figure 3: Data from 14 Stanford Undergraduates (3A), tested in exactly the same way as the children, and 20 Developmental Psychologists who we asked to predict the behavior of the children. As can be seen, while the psychologists were excellent at predicting the behavior of the undergraduates, they failed to predict the behavior of the children on the critical "wug" trial.

Cross-situational Learning: Statistical or Predictive?

These findings have much in common with and are consistent with other cross-situational approaches to word learning (Yu & Smith, 2007; Smith & Yu, 2008), which have established that in word learning tasks, both children and adults can "rapidly learn multiple word-referent pairs by accruing statistical evidence across multiple and individually ambiguous word-scene pairings" (p. 1559). However, in this experiment, we explicitly tested for children's sensitivity to the *information* provided by cues, rather than their co-occurrence rates. This choice was made for two reasons.

First, in many instances, a simple statistical account of word learning cannot effectively rule out the contribution of either innate constraints or other learning strategies, because its predictions overlap to a sufficient degree with markedly different explanations of the same phenomena, such as 'hypothesis testing' (Yu & Smith, 2007 acknowledge this difficulty). While the predictions of a learning theoretic account also overlap with those of high-level constraints such as mutual exclusivity across a wide range of instances, they diverge in certain, critical aspects. Because of this, we were able to test the theories against each other with a highly counterintuitive prediction: that children would choose informativity over mutual exclusivity (or another 'logical' form of inference), even when adults do not.

A second, perhaps more important motivation, was theoretical: we wanted to assess whether children's learning was sensitive to the informativity of cues, and not just simple cumulative statistics. While it is clear that children can and do track conditional probabilities across an array of language learning tasks (Saffran, 2001; Saffran, Aslin, & Newport, 1996; Saffran, Johnson, Aslin, & Newport, 1999; Kirkham, Slemmer & Johnson, 2002), if this were the *extent* of their learning capabilities, they would not be able to master overlapping or context-dependent categories (see Murphy, 2002; Wittgenstein, 1953 for some of the problems inherent to 'real-world' category learning). Our results suggest that children's category learning is informed by competitive, discriminatory processes, which yield markedly different category representations than do non-competitive 'statistical' ones (for reviews, see Smith, Colunga & Yoshida, 2010; Ramscar et al., 2010).

Discussion

The pattern of children's responses indicates that they can and do use informativity in learning to use words. It appears that, as Quine suggested, the words children learn "face the tribunal of sense experience not individually but… as a corporate body." This would suggest that word learning is a systematic, rather than isolated process: what a child learns about any given word is dependent on the information it provides about the environment, in relation to other words (Ramscar et al., 2010). In contrast, it is quite clear that the adults we tested did not place the same value on informativity in their learning that the children did. The adults appeared to reason that if B is not a "DAX" or a "PID," it must logically be a "WUG." Unlike children, it would seem that adults care more about logic than informativity. We should note, however, that while the adult strategy might appear logical in the restricted world provided by our experiment, in the real world, the same object might be "Fido," "a dog," "a dumb mutt" or "pooch" depending upon the context. In this case, the logic of exclusion (Markman, 1990) might not prove to be so helpful, and the strategy adopted by the children may well prove to be a wiser one.

The pattern of data we observed in this experiment further supports the suggestion that young children process information in ways that are qualitatively different to adults (Hudsom Kam & Newport, 2005; Hudson Kam & Newport, 2009), and that this benefits their learning of language (Thompson-Schill, Ramscar & Chrysikou, 2009). The data we report are also consistent with, and may help to illuminate, the many struggles that adult learners of new languages are known to endure (Arnon & Ramscar, 2009). Both of these insights are derived from models of animal learning, in which informativity is a key principle.

Animal models are usually considered irrelevant to language research, and suggesting otherwise can even be seen as undermining human dignity. We demur: although human learning is clearly not identical to animal learning (other animals don't speak), similar objections could be raised in many other areas in which animal models have made valuable contributions to our knowledge. Given that every speaker of every human language on the planet learned the vocabulary that he or she uses, and given that animal models provide our best, most detailed window into the mechanisms that allowed them to do so, there may much insight to be gained by applying animal models to language learning.

Acknowledgments

This material is based on work supported by the National Science Foundation under Grant Nos. 0547775 and 0624345 to Michael Ramscar.

References

Akhtar, N. & Montague, L.J. Lexical acquisition: The role of cross-situational learning. First Language, 19, 347-58, (1999).

Arnon I. & Ramscar M. How order-of-acquisition shapes language learning: the case of grammatical gender, Proceedings of the 31st Meeting of the Cognitive Science Society, Amsterdam, NE, (2009).

Bloom P. How children learn the meanings of words. Cambridge, MA: MIT Press, (2000).

Brown, R. Linguistic determinism and the part of speech. Journal of Abnormal and Social Psychology, 55, 1-5, (1957).

Carey, S. The child as word learner. In M. Halle, J. Bresnan, & G. A. Miller (Eds.), Linguistic theory and psychological reality, Cambridge, MA: MIT Press. 264 –293, (1978).

Chomsky, N. New horizons in the study of language and mind. Cambridge, England: Cambridge University Press, (2000).

Clark, E. V. (1993). The lexicon in acquisition. Cambridge, England: Cambridge University Press.
Colunga, E. & Smith, L. B. From the lexicon to expectations about kinds: A role for associative learning. Psychological Review, 112, 347-382, (2005).
Derks P.L. & Paclisanu M.I. Simple strategies in binary prediction by children and adults, Journal of Exp Psych, 2, 278–285, (1967).
Fodor, J. Concepts: Where cognitive science went wrong. New York: Oxford University Press, (1988).
Gleitman L.R. The structural sources of verb meanings, Lang Acquisition, 1, 3–55 (1990).
Hudson Kam C.L. & Newport E.L. Regularizing unpredictable variation: The roles of adult and child learners in language formation and change, Language Learning and Development, 1, 151-195, (2005).
Hudson Kam, C.L., & Newport, E.L. Getting it right by getting it wrong: When learners change languages. Cognitive Psychology, 59, 30-66, (2009).
Kamin L.J. in: Campbell B, Church R (eds). Punishment and Aversive Behaviour. (Appleton- Century-Crofts, 1969).
Kirkham, N.Z., Slemmer, J.A., & Johnson, S. P. Visual statistical learning in infancy: evidenceof a domain general learning mechanism. Cognition, 83, B35-B42, (2002).
Landau, B., & Gleitman, L. R. Language and experience: Evidence from the blind child. Cambridge,MA: Harvard University Press, (1985).
Markman, E. M. Categorization and naming in children. Cambridge, MA: MIT Press, (1989).
Merriman, W. E. Some reasons for the occurrence and eventual correction of children'snaming errors. Child Development, 57, 942-952, (1986).
Montague P.R., Hyman S.E, & Cohen J.D. Computational roles for dopamine in behavioural control. Nature, 431, 760-767, (2004).
Murphy, G. L. The big book of concepts. Cambridge, MA: MIT Press, (2002).
Naigles, L. Children use syntax to learn verb meanings. Journal of Child Language 17, 357-374, (1990).
Nelson, K. E. Constraints on word learning? Cognitive Development, 3, 221–246, (1988).
Rakison, D.H. & Lupyan, G. Developing object concepts in infancy: An associative learning perspective. Monographs of the Society for Research in Child Development. 73(1): 1-11, (2008).
Ramscar M. & Gitcho N. Developmental change and the nature of learning in childhood. Trends In Cognitive Science, 11, 274-279 (2007).
Ramscar, M. & Yarlett, D. Linguistic self-correction in the absence of feedback: A new approach to the logical problem of language acquisition. Cognitive Science, 31, 927-960, (2007).
Ramscar M., Yarlett D., Dye M., Denny K. & Thorpe, K. The Effects of Feature-Label- Order and their implications for symbolic learning, Cognitive Science, 34, 909-957, (2010).
Rescorla, R.A. Pavlovian Conditioning: It's Not What You Think It Is, American Psychologist, 43, 151-160 (1988).
Rescorla R.A. Probability of shock in the presence and absence of CS in fear conditioning. Journal of Comparative and Physiological Psychology, 66, 1-5, (1968).
Rescorla, R.A. & Wagner, A.R. in Black & Prokasy (Eds.), Classical Conditioning II: Current Research and Theory. (Appleton-Century-Crofts, 1972).
Saffran, J. R. The use of predictive dependencies in language learning. Journal of Memory andLanguage, 44, 493–515, (2001).
Saffran, J. R., Aslin, R. N., & Newport, E. L. Statistical learning by 8-month-old infants. Science, 274,1926–1928, (1996).
Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. Statistical learning of tone sequences byhuman infants and adults. Cognition, 70, 27–52, (1999).
Samejima K., Ueda Y., Doya K. & Kimura M. Representation of Action-Specific Reward Values in the Striatum, Science, 310, 1337–1340, (2005).
Schultz W., Dayan P. & Montague, R. R. A Neural Substrate of Prediction and Reward. Science, 275, (1997).
Shannon C.E. & Weaver W. The Mathematical Theory of Communication (Urbana, Illinois, University of Illinois Press, 1949).
Smith, L.B. Self-organizing processes in learning to learn words: Development is not induction. In C. Nelson (Ed.), The Minnestora Symposia on Child Psychology. Mahwah, NJ: Erlbaum, (1995).
Smith, L. B., & Yu, C. Infants rapidly learn word referent mappings via cross-situational statistics. Cognition, 106, 1558–1568, (2008).
Smith, L. B., Colunga, E. & Yoshida, H. Knowledge as process: Contextually-cued attention and early word learning. Cognitive Science, 34, 1287-1314, (2010).
Quine W.V.O. From a Logical Point of View (Harvard Univ. Press., 1953).
Thompson-Schill S., Ramscar M. & Chrysikou E. Cognition without control: when a little frontal lobe goes a long way, Current Directions in Psych Sci, 8, 259-263, (2009).
Waelti P., Dickinson A. & Schultz W. Dopamine responses comply with basic assumptions of formal learning theory. Nature, 412, 43-48 (2001).
Waxman, S. R., & Gelman, R. Preschoolers' use of superordinate relations in classification. Cognitive Development, 1, 139–156, (1986).
Wittgenstein, L. (1953). Philosophical investigations. Oxford, England: Blackwell.
Woodward, A., Markman, E. M., & Fitzsimmons, C. (1994). Rapid word learning in 13- and 18-month-olds. Developmental Psychology, 30, 553–566.
Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. Psychological Science, 18, 414–420.