Michael Ramscar
Research

My research seeks answers to two core questions: how do human minds make meaningful sense of the world, and how do communities of minds share the meanings they make by way of language?

My goal is to provide answers to these questions in which the principles and mechanisms invoked are described computationally, in as much detail as possible. This allows the explanations proposed in my research to be evaluated using two broad measures: 1. Does the explanation allow clear, testable hypotheses to be derived, and are these predictions matters of empirical discovery? (The idea being that a good model should not only predict unobserved data, it should also predict hitherto unobserved phenomena.) 2. Does the model have practical implications? (While basic research need not lead directly to engineering, my belief is that applications offer a concrete indication that science has made progress.)

In what follows, I describe first the theoretical framework my collaborators and I have developed to answer our core questions, in which language—in both production and comprehension—is treated as a predictive process. I describe the conceptual principles and psychological mechanisms underlying this view, work conducted to test the framework, and some of the discoveries it has led me to. I then describe how we have applied this framework to some well-known, longstanding problems regarding the way children learn language. I also describe other applications of this approach, and the particular conception of cognition it embodies.

How do symbols serve as abstractions?

How do abstract symbol systems work? My work on this question has shown that in learning, the relationship between words and their meanings is asymmetrical, and that the abstract nature of symbol systems imposes some formal constraints on the way that symbolic representation should be conceptualized [1]. This work begins with a analysis of symbolic learning in terms of formal learning theories, which treat learning as a process of acquiring information about cues that *predict* environmental regularities, in which learning arises as the result of differences between what is predicted given a set of cues, and what actually occurs (termed *error-driven learning*). The values of predictive cues are strengthened when outcomes are underpredicted, and weakened when outcomes are overpredicted.

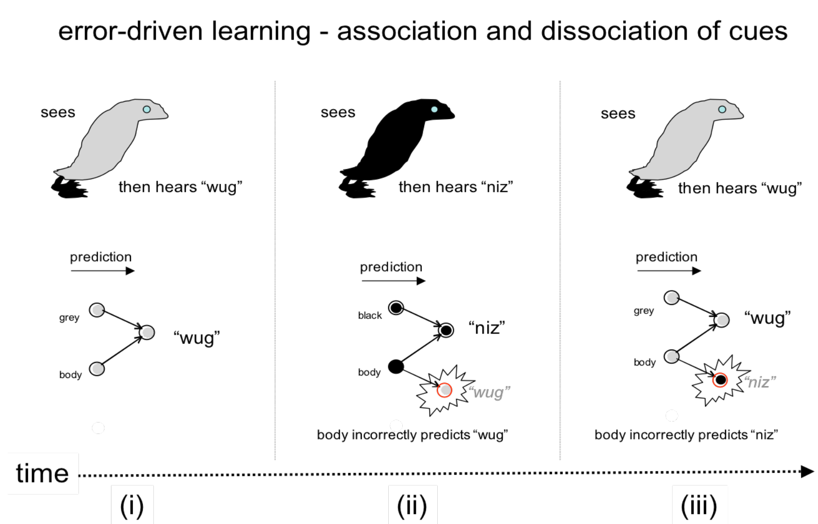

Figure 1. Learning from features to labels. The top panels depict events in the world, and the lower panels cues and the labels in error-driven learning. Note that *body* incorrectly predicts "wug" in (ii). As a result of *cue competition*, the associative strength of *body* to "wug" will be decreased (even though "wug" is not present on this trial); the converse occurs in (iii).

From this perspective, the relationship between labels and their meanings in word learning has two possible forms: labels can predict their meanings; or meanings can predict their labels. Because of the different number of discriminable features that can be discerned in a label, as opposed to what it might represent, the different formal relations that can hold between words and meanings leads to an asymmetry in *what gets learned*, which can be illustrated by considering a world containing two kinds of animal, "wugs" and "nizes". Although both share identical bodies, wugs are grey and nizes are black (Figure 1). To communicate about wugs and nizes, a child must learn to discriminate them based on color.

When features predict labels, this happens easily. Because the feature *body* is a cue to both "wug" and "niz", it causes prediction error: when an animal is present, it predicts "wug" and "niz" regardless of whether the animal is a wug or a niz. These errors get subtracted from the cue value of *body* in learning, causing it to lose out in competition with *black* and *grey*, which don't produce error. The associations between *black* and "wug" and *grey* and "niz" will come to be valued in learning, whereas the associations between *body* and "wug" and *body* and "niz" will be devalued. Although *body* and "wug" co-occur with the exactly the same frequency as *grey* and "wug", learning dissociates the latter pair, enabling the child to learn the set of cues that discriminates "wugs" from "nizes".

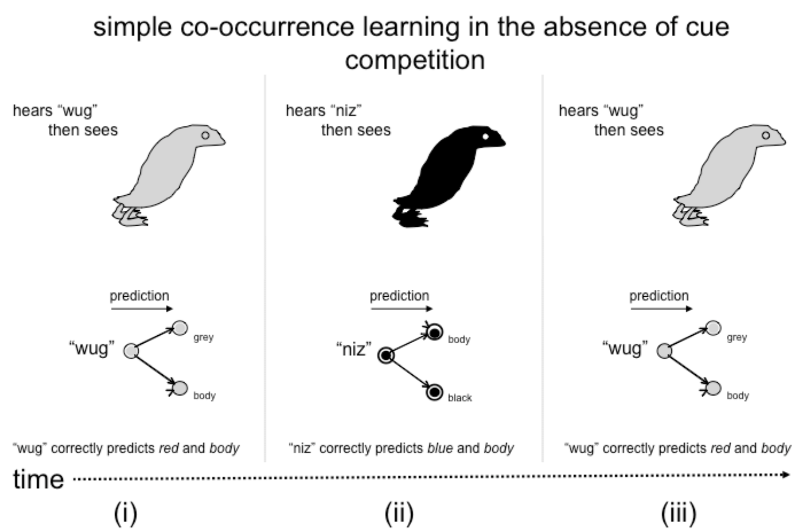

Figure 2. Reversing the order of labels and features eliminates cue competition. No discrimination learning occurs.

Successful discrimination learning arises in Figure 1 because multiple cues compete in learning, and because the non-discriminating feature *body* loses out in this competition. However, when **labels predict objects and their features** in symbolic learning, something different happens. In speech, labels are encountered individually (one word follows another), and the sounds that words are made up of cannot easily be decomposed. Thus when serving as cues, even allowing for variations in intonation, labels provide a would-be learner with few reliably discriminable features that can serve as individual cues. As Figure 2 shows, reducing the number of cues reduces cue competition. Cue competition is necessary for discrimination learning, and in its absence, learning will simply track the rates at which cues and labels co-occur. Error-driven learning thus predicts that different representations will be learned depending on the order in which features and labels are encountered, with better discrimination learning occurring when objects predict labels.

In computational simulations analogous to Figures 1 and 2 above, we manipulated the frequency of "wugs" and "nizes" so that shared features would bias misclassification to the high-frequency category. When an error-driven learning model was trained to predict features from labels (LF learning: Labels predict Features), it failed to learn to discriminate low frequency categories from high frequency competitors with shared features. However, when trained to predict labels from features (termed FL learning: Features predict

Labels), the model learned perfectly. Further, examining the representations learned by the models revealed that the FL-trained model learned about things that *weren't* present. It discriminated well because it downgraded the value of the non-diagnostic cues that erroneously predicted labels that were not subsequently encountered. In the absence of these errors, the LF-trained model simply learned values that reflected the co-occurrence rates of the labels and features [1].

To examine whether these constraints affect human symbolic learning, we asked participants to learn category structures identical to those modeled in the simulations. After training that enforced the two predictive relationships between symbols and exemplars, participants' learning was exactly as our analysis predicted. When trained to predict labels from exemplars, they discriminated and categorized well, but when trained to predict exemplars from labels, they failed to learn to discriminate confusable exemplars [1]. This was despite the fact that in training, each participant saw exactly the same labels and exemplars for exactly the same amount of time; all that varied between conditions was the order in which the labels and exemplars were presented in training.

This analysis offers a solution to a puzzle first noted by Darwin: why do children struggle to learn color words? Our answer to this question begins with the ubiquity of color. In order to dissociate colors from inappropriate labels, children need error. However, given that most of the time children will be exposed to most colors, the lack of systematic covariance between labels and hues will slow discrimination learning drastically. If children's attention could be directed to particular aspects of visual scenes as labels were heard, this process would be simplified. Language can provide children with this direction, but only after they learn names for other things, such as objects (which covary far better with labels; see Figure 1). We conducted a training experiment with two-year-olds that manipulated the order which color words can be used in English (half of the children were trained on "this is a blue ball" while the other half heard, "this ball is blue"). Children who heard object names predict color labels (hearing "this ball is blue" when shown a blue ball) consistently sorted novel objects above chance levels, whereas those who heard color labels predict object names ("this is a blue ball") did not [1]. Another of our studies obtained similar results in number learning.

Abstraction and representation

The work described above shows how symbolic learning is, quite literally, a process of abstraction. Learning to discriminate between categories was facilitated by a process that increases the value of informative cues at expense of value of cues that were uninformative. This process distorts the completeness of what is learned by dissociating non-discriminating features, in much the same way that one might eliminate a great deal of detail to convey the gist of a paper in an abstract. While it is clear that there is much redundant coding in the brain, in follow-up work, we explored whether this process might have an effect on the kind of representations participants formed in learning. We found that when participants were required to categorize objects (discriminating at the category level of abstraction), they performed better when FL-trained. However, when asked to re-identify the exact exemplars they were trained on (discriminate at the exemplar level of abstraction), they performed better when LF-trained [2]. The information gain for the first task had come at the cost of a loss in the second. We suggest that the mechanisms of human learning obey a basic principle: *no representation without taxation*, and from a learning perspective, both costs and benefits arise from learning to conceptualize the environment.

Language as prediction

My work on the formal properties of symbolic learning has led to my development of a conceptualization of communication based on a process of mutual prediction. On this view, when talking, speakers use their cultural and experiential knowledge to generate the utterances they believe are most likely to bring about changes in listeners' beliefs or behavior. At the same time, listeners, far from being passive decoders of tokens of meaning, are using broadly the same process to build up their understanding of what is being said. Listeners use both learned semantic cues to words, and words themselves as cues to other words, in order to predict the behavior and intentions of speakers. Successful communication thus relies on shared prior knowledge that enables mutual predictability, and collaboration between speaker and listener to bring this about [1].

This view of language differs markedly from traditional approaches. Most theories of language are based on the idea of a "sender" and a "receiver" of tokens of meaning. A speaker sends a listener a message in the form of words, which the listener decodes. This natural, folk-view of language suffers numerous problems, which have simply been side-stepped in contemporary linguistic and psychological theories. For example, language is usually seen as referential, such that the relationship between words and their meanings (or the things in the world they represent) is treated as bidirectional. Philosophers such as Wittgenstein and Quine have noted a deep problem in this (what exactly does the word *game* point to?); I have sought to formalize these analyses by considering reference from the point of view of formal theories of information and coding, and learning [1]. My work has shown that reference assumes a theory of reverse abstraction that is incompatible with formal learning and information theories, and the behavior of adults and children in experiments. It shows how the idea that a meaning can be conveyed by a word makes no more sense than the idea that someone might be able to "get" detailed information about the results and method sections of a paper they had never seen, simply by reading its abstract. Given an abstract, one can only make guesses about the results and methods sections, making a kind of prediction about the kind of information they might contain. If the reader is an expert, the likelihood that these predictions will be more accurate, or even substantially correct, will increase. However, *given no more than an abstract*, *the reader can do no more than make predictions*, because the process of abstraction involves discarding information that cannot be later recovered from an abstract representation. Symbols are abstractions, and it follows similarly that symbolic meanings can only be inferred. This is the basis for the idea that language must be a predictive process.

Predicting sequences

Predictive approaches to language have often been argued to be implausible, because the sequential probabilities a learner must acquire have been argued to be unlearnable given the evidence available to a child. These arguments rest in two claims that we believe are extremely weak: that predictions are made only over words, and that there is no generalization in prediction. With Daniel Yarlett, I have developed a detailed cognitive model of cue generalization in predicting word sequences [3; 4], in which predictions about unobserved sequences get smoothed by blending in probability estimates from similar sequences that have been experienced; i.e., knowledge about *p(mat|cat)* is used in estimating the probability of *p(mat|dog)* when *mat* has never previously been heard after *dog*. In simulations trained on data sets that a language learner might reasonably be expected to encounter, our model matches or exceeds the performance of state of the art language engineering models [3]. The model works because it generalizes across neighborhoods of distributionally similar words, and we conducted a training study test the psychological plausibility of this idea. We inserted low frequency words into the distributions of sets of similar words, and then examined the effect of this manipulation on neighbors that were not seen in training. We found that inserting a low frequency word like *samovar* into the distributions of a set of words that were in turn distributionally similar to, say, *kettle*, resulted in measurable changes in priming between *kettle* and *samovar,* and to semantic similarity judgments between them. Since participants were never exposed to *kettle* in training, these results support the idea that linguistic experience does indeed generalize across distributional neighborhoods [4].

In conjunction with this work, Colin Bannard, I have examined the *kinds* of representations that people learn probabilities for. In a series of experiments, we have found that that children and adults learning of probabilistic information about language is not restricted to words and phonemes; their sensitivity to manipulations of sequence frequency in which word and bigram frequencies are held constant reveals that they also acquire a detailed knowledge of the probabilities of multi-word sequences [5, see also 6].

Learning words in context

It is clear that children do not just face the task of mapping sounds and meanings; they also have to learn to identify patterns of regularity in the sounds themselves. In our models, learning is treated as a process of iterative discrimination, in which the cues to individual linguistic regularities are slowly refined, along with those regularities themselves. Accordingly, since words will generally be encountered in linguistic contexts, both meaning and preceding words serve as potential cues to later words. To explore the dynamics of this kind of learning in context, we have used our predictive learning model as the basis for an investigation of the reasons behind adults' poor performance in learning grammatical gender during second language acquisition. There are clear differences in the way that adults and children approach the task of language learning: unlike children, adults already have a language, as well as folk theories about how language works, and a greater proportion of their initial learning may take place in a classroom, as opposed to in context. If adults learn the meanings of words out of context, our model suggested that this would have the effect of blocking later learning of associations between linguistic contexts and words. To test this idea, Inbal Arnon and I trained participants on an artificial language, utilizing a classic blocking design: half of our participants were trained on blocks of noun–meaning mappings prior to encountering the nouns in sentences, along with their meanings, whereas for the other half this order of training was reversed. Consistent with our blocking hypothesis, we found that the participants who first encountered the nouns in sentences learned to map them to their gendered determiners far better than those who first trained on the nouns. Moreover, the sentence-first group also learned the noun-meaning mappings better, suggesting that grammatical gender markers may serve to carry information by narrowing the scope of subsequent lexical predictions [7].

In related work, my collaborators and I have found the asymmetry we identified in our studies of Feature-Label-Order is consistent with a bias across languages for inflections to be added to word endings. A corpus analysis of English confirmed the prediction that suffixes are more informative about the grammatical category of root-words than prefixes, while an artificial language learning task revealed that suffixes (which are predicted by root-words, i.e., FL-learning) were learned significantly more accurately than prefixes (which predict root-words, i.e., LF-learning) [8].

A surprising prediction about over-regularized plurals

We have used the framework described above to reanalyze the reasons behind children's tendency to over-regularize plurals (i.e. *mouses*), a key question in debates about the nature of language and learning. In our predictive model of learning, over-regularization arises initially because the semantic cue of 'plurality in general' can cue both regular and irregular forms [9,10]. This causes interference and over-regularization, because the frequency of regular plurals in English causes their representations to strengthen more rapidly than those of irregular forms. In the model, this interference resolves itself naturally in time, as a result of error-driven learning. Until 'plurality in general' has been unlearned as a cue, the occurrence of regular plurals leads to irregulars being expected, which in turn generates negative learning when regular plurals are encountered. In time, this negative learning dissociates the irregulars from the more general plural features that lead to interference and over-regularization. The model makes some very strong predictions: that children can stop over-regularizing simply by being exposed to a normal distribution of plurals, they will do so without the need for explicit feedback, and even when repeating the plurals they over-regularize [10]. Moreover, it predicts a point in learning where exposing children to *only* regular plurals will *reduce* over-regularization [4]. Studies with children have provided strong support for all of these predictions [10, 11].

Operationalizing "use" in meaning

I have also used our predictive framework to develop a novel account of the metaphorical priming effects my collaborators and I had uncovered in a series of studies [12, 13]. In these experiments, we found that getting people to think about abstract forms of motion in turn affected the way that they answered ambiguous temporal questions, however the underlying cause of these effects was unclear. In a recent series of experiments, we examined whether modeling the expectation produced by the words in the primes in our experiments could be responsible for the effects we observed. In an earlier study, we had shows how manipulating an integer (say four, or twenty) in an otherwise identical sentence could result in different temporal priming effects. Accordingly, we examined the distribution of words following—cued by—a given integer, to see if they might in turn cue words with more or less future bias. This simple model proved to be a really good predictor of how people respond to the ambiguous question. The more an integer (say *eleven*), cues time words (*weeks, hours*; which in turn prime future words like *later*), the more people believed that a meeting that had "been moved forwards two days" was now on "Friday" [14].

Number words prime future interpretations because, as the distributions of English reveal, we tend to talk about the future with a greater level of detail than the past, such that in the distribution of English, once one has encountered "eight minutes", "later" is far more likely than "earlier"). Not only do our results suggest that people are sensitive to these semantic biases in language distributions, but they also point to the way in which a language can serve as a repository of cultural knowledge, a factor that is often ignored in psychological theorizing. Although people can obviously learn about the different levels of detail appropriate to describing future and past events by other means, distributions provide a medium through which children can implicitly learn many aspects of their cultures, and the particular views of the world enshrined therein (thereby reducing the degree to which one has to assume children form their understandings of the world based on individual experience alone).

The architecture of learning as an adaptation for culture

A key question that all theories of language must answer is this: what are the origins of linguistic conventions? Symbolic communication is, in its essence, conventional. Given a symbol, a social animal needs to be able to infer and understand (and often, to do) the *appropriate thing* in the *appropriate context*. For this to happen, "symbolic values," must be conventionalized and internalized. We suggest that part of the answer to this question lies in the slow pattern of development of control processes in prefrontal cortex (PFC) in childhood [15, 16]. PFC functions enable adults (and animals) to filter their behavior and attention, and thus direct their learning. A great deal of evidence – behavioral, computational and neurobiological – supports the idea that children barely select between, or filter competing responses at all prior to their fourth year, and that adult levels of PFC function develop very slowly, over the course of childhood. We suggest that since PFC function provides a mechanism for filtering, it will result in more "individualized" patterns of learning, whereas the absence of PFC function (and filtering) will result in learning that is more environmentally determined and conventionalized. In learning the appropriate cues to symbols, unsupervised cue competition will tend to produce very similar patterns of learning if learners are exposed to similar distributions of environmental cues and symbols. From this perspective, the delayed pattern of PFC development can be seen as an adaptation for cultural and linguistic convention learning.

Other work and future directions

The view of language described here differs markedly from most other contemporary theories, which adhere to a view of language built around senders and receivers of words with discrete meanings. The field has largely dodged the problem of explaining how discrete lexical concepts are acquired, how they contribute meaning to a compositional syntax, and the many seemingly intractable problems associated with this. (Recognition for these problems is apparent in Chomsky's and Fodor's arguments that even concepts like *bureaucrat* are innate. Conceptual nativism doesn't, however, *solve* any of these problems, since to date no mechanistic account of the development, structure, or computational properties of innate concepts has been offered; claims about innate concepts are statements of faith in a particular view of language, not theories.)

By taking learning as a starting point, we have developed an approach to language consistent with the idea that people learn how symbols are *used* in communication, and not discrete word meanings. Theoretically, this approach is consistent with the ideas of major linguistic philosophers such as Quine and Wittgenstein. Methodologically, treating communication as prediction offers traction when it comes to understanding the learning, processing and nature of language. Several ongoing lines of work seek to both apply our predictive approach to an even greater range of communicative phenomena, and to continue to develop our theoretical understanding of the predictive basis of communication.

An equally attractive aspect of this approach to me is practical: it has enabled me to build models that successfully predict hitherto unforeseen phenomena in language learning, and to formally design training experiments that have succeeded in teaching color word discrimination at a much earlier age than had been thought possible. As well as further developing the account of learning in children described here – in particular, we want to better understand the relationship between explicit and implicit learning – we see the development of methods to apply these ideas to domains such as early word learning as an important future direction for our research. In one on-going line of research we are examining whether the development of PFC function interacts with children's response to different kinds of explicit feedback, while in another we are developing training media for infants and young children based on our theoretical approach, and conducting experiments to test their efficacy.

Publications and Manuscripts Cited

[1] Ramscar, M., Yarlett, D., Dye, M. Denny, K.G., & Thorpe. K.L. (in press). Feature-Label-Order effects and their implications for symbolic learning. *Cognitive Science*.

[2] Ramscar, M. & Dye, M. (2009b). No representation without taxation: The costs and benefits of learning to conceptualize the environment. *Proceedings of the Analogy 09 Conference*, Sofia, Bulgaria.

[3] Yarlett, D (2008) Language Learning Through Similarity-Based Generalization, PhD Thesis, Stanford University.

[4] Yarlett, D., Ramscar, M. & Dye, M. (in submission) Language Learning Through Similarity-Based Generalization. In second round review at *Cognitive Science*.

[5] Bannard, C. (2006) Acquiring Phrasal Lexicons from Corpora. PhD Thesis, University of Edinburgh.

[6] Bannard, C. & Matthews, D. E. (2008) Stored Word Sequences in Language Learning: The Effect of Familarity on Children's Repetition of Four-Word Combinations, *Psychological Science*, 19, 241-248.

[7] Arnon, I., & Ramscar, M. (2009) Granularity and the acquisition of grammatical gender: How order-of-acquisition affects what gets learned. *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, Amsterdam, Netherlands.

[8] St. Clair, M., Monahan, P., & Ramscar, M. (2009) Relationships between Language Structure and Language Learning: The Suffixing Preference and Grammatical Categorization. *Cognitive Science*. 33(7), 1317-1329.

[9] Ramscar, M.(2002) The role of meaning in inflection: Why the past tense does not require a rule. *Cognitive Psychology*, 45(2), 45-94.

[10] 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.

[11] Ramscar, M., & Dye, M. (2009) Expectation and negative evidence in language learning: the curious absence of mouses in adult speech. *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, Amsterdam, Netherlands. A manuscript describing an extended version of this work is under review at *Nature*.

[12] Boroditsky, L., & Ramscar, M. (2002). The roles of body and mind in abstract thought. *Psychological Science*, 13, 185-188.

[13] Matlock, T, Ramscar, M., & Boroditsky, L. (2005). On the experiential link between spatial and temporal language. *Cognitive Science*, 29, 655-664.

[14] Ramscar, M., Matlock, T., & Dye, M. (in press) Running down the clock: the role of expectation in our understanding of time and motion. *Language and Cognitive Processes*.

[15] Ramscar, M., & Gitcho, N. (2007). Developmental change and the nature of learning in childhood. *Trends In Cognitive Science, 11*(7), 274-279.

[16] Thompson-Schill, S. L., Ramscar, M., & Evangelia, C. G., (2009) Cognition without control: When a little frontal lobe goes a long way. *Current Directions in Psychological Science*. 18(5).