Michael Ramscar, Melody Dye, James Blevins and Harald Baayen

9 Morphological development

1 Introduction

Language, whether spoken or written, is the primary means by which humans communicate. Yet exactly how communication is achieved through language has yet to be explained. This chapter describes our approach to understanding how morphological systems work and what morphological development entails. The approach is rooted in the way minds learn, and is based on clearly and explicitly stated learning mechanisms for which there is a wealth of biological evidence; It is consistent with the principles that govern artificial communication systems (Shannon, 1948); And, unusually, it makes surprising – and successful – predictions about the pattern of morphological development. Because it is rooted in the discriminative principles of learning, the perspective on language our approach offers is very different to the associative, combinatoric view taken by most researchers. However, we believe that this perspective will prove to be of fundamental importance to understanding human communication, and consequently, the challenges facing children with communicative disorders. Accordingly, in this chapter we describe the principles of learning in detail, along with the picture of morphological and linguistic development they give rise to.

2 Language, morphology and development

In thinking about how human communication works, linguists have typically assumed that language facilitates communication by conveying meanings, much as trains convey passengers (Reddy, 1979). Linguistic theories assume that words – and the sub-word units called morphemes – encode units of meaning, such that the word 'units' is composed of two morphemes (the morpheme 'unit' and the morpheme 's'), and the word 'morphemes' is composed of three ('morph,' 'eme' and 's').

From this perspective, the task facing a language learner can be broken down into a three-fold process. It involves learning what the conceptual units of her language are, learning how to associate them with sound units to create morphemes, and figuring out the kinds of morphemes that can be combined to form complex words (along with how sequences of morphemic combinations combine to yield higher-order sequences, such as sentences).

Michael Ramscar and Harald Baayen, Eberhard Karls Universität Tübingen
Melody Dye, University of California at Berkeley
James Blevins, University of Cambridge

DOI 10.1515/9781614514909-010

Brought to you by | De Gruyter / TCS
Authenticated
Download Date | 7/3/18 7:37 PM

Over a century of study has uncovered a number of problems with this approach. Critically, at both a behavioral and neural level, it has been found that learning simply doesn't work in the associative manner that linguistic theories are wont to imagine. What has traditionally been called "associative" learning is not associative at all. Rather, it is a systematic process that serves to discriminate the details of a learner's internal representation of the world (Ramscar et al, 2010).

Further, while humans are perfectly capable of learning to discriminate between events and behaviors, they do so in ways that do not involve the discrete internal representations (i.e., the "units of meaning") that are supposed to provide the stock of combinable elements in combinatorial approaches to language (Wittgenstein, 1953; Ramscar and Port, 2015). Finally, though by no means exhaustively, combinatorial approaches to language describe meanings as being encoded into verbal and written signals by speakers and writers to later be decoded by readers and listeners; However, the process this envisages violates the basic principles of coding theory, as it assumes that the appropriate meanings of words or morphemes with many potential "senses" can be successfully decoded from signals that do not actually code for these senses. (In the case of written English, the available coding resources do not even consistently discriminate different lexical forms from one another, such as the past and present tenses of *read*).

Although most cognitive scientists are familiar with at least some of these problems, most theories of language acquisition and processing remain steadfastly rooted in associative combinatorics, even in the absence of an adequate account of what exactly gets combined, or how the encoding and decoding of meanings is actually supposed to work. In what follows, we describe an account of language learning and processing based on *discrimination learning*. It does not assume that morphemes serve to convey meanings, but rather that they serve to discriminate between meaningful states of affairs in communication. From this perspective, the task facing a language learner is that of learning which potential states of the world morphemes may discriminate, and how the distinctions afforded by morphemes can be used in communication. To illustrate the utility of this approach, we show how it not only provides a satisfactory account of many phenomena associated with morphological development, but also makes surprising (and successful) predictions about the way that patterns of morphological over-generalization develop and recede.

3 How do children learn?

Any account of how children master morphological processing seeks to answer two questions: *what* do children learn, and *how* do they learn it? In the recent past, a great deal of linguistic theorizing has proceeded from *what* to *how*. Perhaps unsurprisingly, these theories have struggled to explain how the *what* of their

Brought to you by | De Gruyter / TCS
Authenticated
Download Date | 7/3/18 7:37 PM

theoretical postulates are learned. This has, in turn, led to a situation where the psychological bases of many theories of language are opaque, and where claims about "innate mechanisms" abound.

We take the opposite tack: starting from *how* children learn, we consider how this might constrain our conjectures about *what* children learn. One potential benefit of this approach is that our current scientific understanding of learning is far more advanced than that of language. We share many of our basic learning mechanisms with other animals, and in many domains, animal models have proved invaluable in illuminating the biological and neural structures underlying these mechanisms.

Legend has it that Ivan Pavlov's famous discovery – that ringing a bell before giving dogs food later caused the dogs to salivate whenever the bell sounded – was a felicitous accident.[1] Whatever the exact truth of the matter, his discoveries have given rise to a popular idea: That animals learn to "associate" unrelated events according to the frequency with which a stimulus (a bell) and a response (salivation) are paired. Empirically, however, this naïve view of conditioning – as a process that simply tracks co-occurrences – has long been known to be wrong (Rescorla 1988), as have two popular (yet equally false) beliefs about the conditions that produce learning: First, that explicit rewards and punishments are necessary for learning; and second that a co-occurrence between a stimulus and a response is sufficient for learning.

Empirically, mere association cannot account for conditioning. For example, if a group of rats is trained on a schedule of tones and shocks, they will quickly learn to associate the tone and shock, and freeze when later tones sound. However, if rats are exposed to an identical number of tone-shock pairings into which a number of tones that are *not* followed by shocks are interpolated (i.e., *Tone, Tone, Tone-Shock, Tone-Shock, Tone, Tone, Tone, Tone, Tone-Shock, Tone, Tone, Tone, Tone, Tone-Shock… etc*), the rats show a different pattern of learning. As the number of tones not followed by shocks increases, tone-shock associations decrease proportionally (Rescorla, 1968).

Given that the *non*-occurrence of shocks after tones influences the degree to which rats condition to the tones, it follows that learning must comprise more than simple counts of positive co-occurrences of cues with events. To explain how rats learn from the *background rate* of the tones (a phenomena that popular "associative" conceptions of learning cannot explain; Rescorla, 1988), modern theories of learning suppose that in the two situations just described, the "no shock" trials act to alter rats' implicit expectations about the tones. These theories conceive of learning as a process that serves to reduce uncertainty in the predictive mental models that learners construct out of their experience with their environments (Rescorla, 1988). Functionally, learning modulates the value that an organism implicitly assigns to sensory cues as predictors of the events that it experiences. As events unfold, representations of cue values change as a function both of current expectations, and the degree to which

1 The bell is actually a myth (Todes, 2014).

Brought to you by | De Gruyter / TCS
Authenticated
Download Date | 7/3/18 7:37 PM

these expectations have led the learner to anticipate (or fail to anticipate) what actually occurs.

Importantly, learning to predict an outcome from one cue has an impact on the uncertainty that drives the learning of other cues. This is best illustrated by *blocking* (Kamin, 1969): If a rat learns it will be shocked upon hearing a tone, and then later a light is paired with the tone, learning about the light as a cue to the shock will be inhibited, because the tone is already fully predictive of the shock. Thus the rat will be unlikely to freeze in response to the light alone.

These and other results demonstrate that rats do not learn simple "associations" between stimuli and responses, but rather learn the degree to which cues are systematically informative about events, a process that discriminates cues that are more informative from cues that are less so (Rescorla, 1988). Since there are invariably far more uninformative coincidences in the environment than informative ones, it follows that expectations that are wrong have more influence on the shape of learning than expectations that are right – which is why discrimination learning is often described as being *error-driven*.

Finally, we should stress that although our rat learning example focused on tones, in principle, *everything* in the local environment could have potentially influenced learning (Rescorla, 1988). However, in the same way that rats learn to discount tones as predictive cues the more they encounter them absent shocks, they also learn to discount other aspects of their environment that have high background rates relative to a relevant outcome. For the sake of simplicity, models and explanations tend to focus solely on potentially informative cues, ignoring cues whose high background rates are likely to render them largely irrelevant in competitive terms. In principle, however, the novelty of a cue is entirely relative, and can only be computed in relation to the other available cues, and a learner's previous experience with them (Rescorla 1988).

Beginning with Rescorla and Wagner (1972), many formal models employing these principles have been devised to fit and predict learning effects. The Rescorla-Wagner learning rule is a discrepancy function that takes the difference between expectation and reality on a given trial, and uses this to update expectations by modifying the values of a set of cues in relation to that outcome. Stated formally, learning occurs whenever the outcome of a given trial fails to match the expectations generated by the available environmental cues. Cue values V_X are updated in proportion to the overall mismatch, which is given by the discrepancy function ΔV. That function takes the difference between the outcome λ and the summed predictive value of all cues present, V_{tot}, and weights it by two parameters, which denote cue salience (α), and learning rate (β).

$$\Delta V_X^{n+1} = \alpha_X \beta (\lambda - V_{tot})$$

Brought to you by | De Gruyter / TCS
Authenticated
Download Date | 7/3/18 7:37 PM

Formally, λ represents the maximum predictive value for an outcome. On trials where the outcome occurs, λ is usually set to 1, and on trials when it does not, to 0. The configural value of the cue set is updated as follows:

$$V_{tot} = V_X^n + \Delta V_X^{n+1}$$

As this suggests, learning tracks predictive accuracy. Cue values are strengthened when an outcome's likelihood is underestimated, and weakened when the likelihood is overestimated. Learning is also a zero-sum game. The predictive value 'lost' by one cue, can be subsumed by other cues, leading to competition between cues, and preferential strengthening of the most reliable ones, and discriminatory weakening of others. The trajectory of this competition is shaped both by 'positive evidence' (co-occurrences between cues and predicted events) and 'negative evidence' (the absence of a predicted event following a cue).

While much of the impetus for the development of this kind of learning rule came from behavioral experiments in animals, there is now good evidence for their neurobiological basis in humans (see Schultz, 2006). Learning rules of this type accurately predict patterns of synaptic firing in midbrain dopamine neurons in learning tasks (Waelti, Dickinson, and Schultz, 2001) and have been productively applied to many aspects of human behavior and cognition, such as decision making, executive function, habitual learning, and response selection (Montague, Hyman, and Cohen, 2004), demonstrating considerable predictive and explanatory power.

While for historical reasons the Rescorla-Wagner learning rule is often characterized in associative terms (see e.g. Miller, Barnet, and Grahame 1995, Siegel and Allan 1996), it is important to note that in computational terms, the rule describes a *discriminative* learning mechanism (Ramscar et al, 2010; Ramscar et al, 2013c). Since this learning rule is arguably the best-supported formalism in psychology, with a sizeable – and still growing – body of behavioral evidence arguing in support of its principles (Miller, Barnet, and Grahame 1995; Siegel and Allan 1996; Ramscar, Dye, and McCauley 2013), characterizing learning in discriminative terms can help elucidate the mechanisms that govern neuropsychological development (Ramscar et al, 2010; Ramscar, Dye, and McCauley 2013).

Indeed, adopting this approach has proven useful to both predicting and understanding morphological development as children learn language. Experimental results offer reason to believe that children are exquisitely attuned to the structure of their linguistic environments in precisely the manner predicted by discriminative models (Ramscar et al. 2010, 2011, Ramscar at al 2013a).[2] Indeed, in our research, we have found that these models successfully predict patterns of development in set-size

2 Although Saffran et al (1996) claim that their "statistical learning" results are incompatible with Rescorla-Wagner learning, given a flat cue structure – like a stream of syllables – the model actually asymptotes at cue weights that approximate the transitional probability between each syllable (Ramscar et al, 2010), which behaviorally is what Saffran et al observed in infants.

Brought to you by | De Gruyter / TCS
Authenticated
Download Date | 7/3/18 7:37 PM

learning (*subitization*; Ramscar et al., 2011) and rule-understanding in children (Ramscar et al., 2013), as well as changing patterns of performance in simple lexical "association" tasks across the lifespan (Ramscar et al., 2013; Ramscar et al., 2014).

Given this, and given our aim of using *how* learning works to uncover *what* is learned in morphological development, it is worth noting that the logic of discrimination suggests that, far from the "blooming, buzzing confusion" of multiple entities described in much of developmental psychology, a newborn learner is best conceptualized as entering the world with a large, undifferentiated set of cues connected to little or no environmental knowledge. Starting from N = 1, the set of (more or less individuated) entities in the learner's representation of the world then begins to expand as perceptible variances and invariances in the environment encourage discrimination learning (James, 1890).

4 What is morphology, and how might it be learned?

4.1 The combinatoric approach

Having established in broad terms *how* children learn, we now turn our attention to *what* they learn. That is, we can now consider whether discriminative learning mechanisms are sufficient to account for morphological development, and if so, how.

We ought to acknowledge here that interest in morphological development extends far beyond the simple concern of understanding, say, plural marking, for its own sake. Over the course of the past quarter century, research on morphological development has been seen "as addressing some of the most important issues in cognitive science" (Seidenberg and Plaut, 2014, p. 1), largely because:

> [Morphology has…] three interesting characteristics. First, it is systematic: Most past tenses are formed by adding the morpheme that is spelled -ed and pronounced as in the examples baked, baited, and bared. Second, it is productive: People can readily generate past tenses for novel forms such as nust-nusted or wug-wugged. Third, it is quasiregular (Seidenberg and McClelland, 1989): There is a main pattern but also irregular forms that deviate from it in differing degrees (e.g., keep-kept, run-ran, go-went). Phenomena such as tense on verbs and number on nouns have been taken as simple, decisive demonstrations that grammatical rules are an essential component of linguistic knowledge (Pinker, 1999). Irregular forms exist outside this system of core linguistic knowledge and are learned and generated by other mechanisms such as memorization and association.

Rumelhart and McClelland's (1986) model offered an alternative view of this last point. Taking the phonological form of a verb's present tense as input, it generated the phonological form of its past tense as output using a uniform procedure for all tenses. It also supported the generation of past tense forms for novel verbs. The

Brought to you by | De Gruyter / TCS
Authenticated
Download Date | 7/3/18 7:37 PM

model and the various claims made about it caused controversy and launched a debate that has generated an enormous body of research on a range of morphological phenomena.

For our current purposes, however, the agreements in this debate are more relevant than its disagreements: Almost all the participants in this debate accept that the *what* of morphological development is a means of composing and decomposing associative morphemes. Hence, in the case of English plurals, they assume that a child learns a morpheme that associates the concept *mouse* with the word "mouse", an association between the concept *mice* and "mice", an association between the concept *rat* and "rat", and an association between the concept for *sets of objects* (excluding multiple mouses) and a morphemic gesture characterized linguistically as a terminal sibilant on a noun, and often written as +*s*, etc.

Yet, as we described above, the *what* of the "associative" learning process has actually been found to be discriminative. The "association-tracking" at the heart of the processes imagined in this debate is actually an evolutionary twist on Sherlock Holmes' dictum – "When you have eliminated the impossible, whatever remains, however improbable, must be the truth." When a rat learns to associate a tone with a shock, the "association" is what is left over after learning has systematically weighed every other potential source of information and found it wanting.

Consistent with this picture of learning, researchers studying human categorization have found that at a behavioral and neurological level, human performance is best accounted for by models that don't actually contain pre-established (or even determinable) categories, but instead treat categorization as an active process of discriminating between more or less appropriate category labels (or other affordances and behaviors) in context (see Ramscar and Port, 2015 for a review). Similarly, the non-discrete nature of meanings could be described as the closest thing that philosophy has to offer to a fact (Wittgenstein, 1953; Quine, 1960). All of which indicates that the idea of the "associative morpheme" – a discrete mapping between a discrete unit of meaning and a discrete linguistic unit – is incompatible with what we know about human learning; about the nature of meaning; and about the computational properties of the human categorization system.

What is more, if we accept the message of this evidence, and allow that morphological systems do not involve a discrete set of mappings between a set of units of meaning and the lexical forms of a language, then a significant peculiarity of this debate – and indeed, of the way linguists think about morphology more generally – becomes apparent. In the Rumelhart and McClelland (1986) model and most subsequent models of morphological development, the child is envisioned as learning a combinatorial, transformational rule that (e.g.) adds the English past tense morpheme +ed to a verb stem in order to transform an uninflected form into a past tense form. Accordingly, the model's training set was devised to teach this transformation; uninflected forms are repeatedly turned into past tense forms, as if the learning

Brought to you by | De Gruyter / TCS
Authenticated
Download Date | 7/3/18 7:37 PM

environment comprised mature speakers going about saying, "*go-went, walk-walked, speak-spoke, talk-talked, etc.*"[3]

However, this characterization of the learning environment is dubious in the extreme. Mature speakers do not wander around giving extensive tutorials on the nature of supposed transformations, saying, "*go-went, walk-walked, etc.*" They talk about what interests them, in context. This means that children rarely, if ever, hear "*go*" then "*went*" in close proximity. Instead, they hear "Wanna *go* see a movie?" on Tuesday, and "Oh no! I forgot to put the trash out before we *went* to the movies last night," on Wednesday.

In other words: 1) Children don't learn in the way envisaged by the combinatoric account of morphology; 2) The units it supposes that they learn are highly implausible; And 3) Children learn morphological systems without ever actually encountering morphemes being used transformatively. Thus it seems possible, perhaps even likely, that a different approach to the conceptualization of morphological processing might yield a better account of *what* develops, as well as *how*.

4.2 The counter-intuitive appeal of discriminative morphology

We begin our formulation of this alternative account by considering the very things that the combinatoric story gets wrong: the nature of learning, human knowledge representation, and the learning environment. As we noted above: learning is discriminative; categorization is an active, context driven process that serves to discriminates between lexomes (and other more or less discrete affordances and behaviors); and children hear lexomes used in contexts that offer precious little evidence for transformations at all.

To illustrate how these constraints might influence what a child actually learns, we take English nominal morphology as an illustrative example, and consider how these factors might give rise to the typical patterns of over-regularization children exhibit while learning the nominal system. A fuller description of this approach, along with code to produce the simulations described below, can be found in Ramscar et al. (2013d).

In English, correct irregular plural marking is particularly difficult to acquire (Ramscar and Dye, 2011), even in comparison to the more commonly studied case of past tense marking. This reflects the nature of the input. Consider verbs: While irregulars are rare as types, they tend to have high token frequencies, such that the 40 most frequent verb forms are all irregular (Davies, 2009). Moreover, in the Reuters

3 Although Rumelhart and McClelland (1986) use a discriminative learning rule analogous to Rescorla-Wagner, they describe their model as a pattern associator, and adopt a combinatoric view of morphology.

Brought to you by | De Gruyter / TCS
Authenticated
Download Date | 7/3/18 7:37 PM

Corpus (Rose et al., 2002) just three irregular verbs (*be*, *have*, and *do*) account for fully a quarter of the attested verbs forms, with past tense verb forms outnumbering base or present tense verb forms. Thus, in learning past tense inflection, children are likely to encounter more past inflected than uninflected verb forms, and of these, more irregular than regular past inflections.

Noun plurals are different: Children mainly encounter singular forms, and when they do encounter plurals, they are likely to be regular. In the Reuters Corpus, only around 30% of nouns occur in their plural form, and of these, the overwhelming majority of types and tokens are regular. While this makes the learning problem for plurals substantively more difficult than the past tense, the two problems may not be different in kind; As with the past tense, children's irregular plural production follows a 'U-shaped' developmental trajectory, such that children who have been observed to produce "mice" in one context still frequently produce over-regularized forms such as "mouses" in another (Arnon and Clark, 2011).

Our account of nominal development assumes, first and foremost, that native languages are not learned in formal teacher-pupil settings, but by hearing them used in context. We also assume that lexical and morphological systems are systematically informative, such that the combined value of both positive and negative evidence favor a mapping between the experiences a child has with the class of objects described by a given noun, and the noun itself. For example, a child learning the lexome (or lexical form) "mice", will hear the word used in a way that makes it most informative about mice, or depictions of them, and must learn to associate the appropriate cues in the environment (*mouse-things*) with that lexome (Quine, 1960, Wittgenstein, 1953). Conceptually, this assumption reflects the idea that adult speakers use language in informative ways, and hence, that a *mouse* ought to be more informative about the lexome "mouse", and *mice* more informative about the lexome "mice", than vice versa (otherwise, "mice" could equally mean *mouse*, and "mouse" could equally mean *mice*). Accordingly, when a child is asked to name a picture of *mice*, the child is able to say "mice", because learning has discriminated a set of mappings between the mice-relevant semantic dimensions of the child's experience and the more or less discrete gestural/phonetic form "mice".

At the outset of word learning, all and any kind of *stuff* in the world will seem potentially informative about concrete nouns such as "mouse" and "mice". Thus learning to discriminate the correct cues to each form will involve discriminating the *particular stuff* that is best associated with any given singular and plural form of a noun (e.g., *mousiness* in the case of "mouse" and "mice") from other kinds of stuff. At the same time, learning to discriminate singulars from plurals will require discriminating the specific stuff that best predicts one as opposed to the other (i.e. the presence of *multiple mouse objects* ought to predict "mice", as opposed to a *single mouse object*, which ought to predict "mouse").

Brought to you by | De Gruyter / TCS
Authenticated
Download Date | 7/3/18 7:37 PM

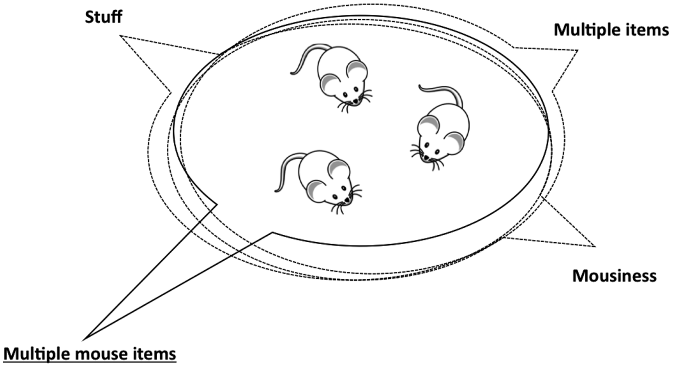

Figure 1: Four semantic-dimensions that will be consistently reinforced by a child's exposure to the word "mice" in the context of mice. (While these dimensions are separated for explanatory clarity, as far as the learning mechanisms are concerned, they could equally be values on continuous perceptual dimensions)

To illustrate how these distributional differences influence the learning of English nominal morphology, Figure 1 depicts four environmental dimensions that will reliably and consistently covary with the irregular plural form "mice." Although these semantic dimensions will co-occur with the word "mice" at the same rate, their covariance with other singular and plural nouns will differ. Because of these differences in background rate, the error associated with each dimension will vary in kind. Accordingly, although early in learning, generic cues like *stuff* will receive positive reinforcement when "mice" is encountered in context, their ubiquity will also cause them to produce a high degree of error. Indeed, compared to more uniquely informative cues, they will occur far more often in contexts in which "mice" is not heard. Thus, the influence of the more generic cues will wane over the course of learning, as learners converge on *multiple mouse-items* as the best cue to "mice."

In learning, error arises as a function of experience, which means that the pattern of reinforcement and unlearning of the semantic dimensions in Figure 1 depends heavily on their distribution in the learning environment. While the set of singular and plural lexomes that are usually classed together and called 'regular' plurals in English are distinguished by a number of subtle differences (such as different sibilant allomorphs), broadly speaking, regular plurals are far less discriminable from one another than irregular plurals, particularly in terms of how they discriminate plurality from singularity. In regular plural lexomes, plurality is uniformly denoted by the presence of some form of final sibilant. By contrast, irregulars employ a variety of means of discriminating singular and plural forms. This guarantees that irregular plural lexomes are, at once, less similar to their singular forms than regular lexomes (e.g., "dog" / "dogs" vs. "child" / "children"), and are also less similar to other irregulars ("foot" / "feet" vs. "child" / "children").

Brought to you by | De Gruyter / TCS
Authenticated
Download Date | 7/3/18 7:37 PM

At this point, it is worth expanding a little on what we mean by a lexome: In our model of nominal learning, the challenge is seen as one of learning to discriminate the semantic (and lexical) cues to a system of phonetic and lexical contrasts simultaneously in context (Ramscar, Dye, and Klein, 2013; Ramscar and Port, 2015). As such, the degree to which any given phonetic lexical contrast has itself been discriminated will depend entirely on the current state of the learner.

Having said this, the discrete lexomes described here represent a simplification for descriptive and modeling purposes of what is, in reality, a far more continuous system. The requirements of the modeling task will dictate the specific representations adopted. For instance, it might be helpful for some theoretical purposes to represent the state of the plural system mastered by a young learner as comprising irregular forms, regular stems, and a single lexome marking the regular plural contrast +*s*, whereas for other purposes, the regular plural contrast might be more appropriately represented by positing different lexomes for each regular plural allomorph. For example, in modeling the adult ability to discriminate and respond to prosodic differences in the stems of regular forms (Baayen et al., 2003, Kemps et al., 2005), well learned regular plural and singular forms might best be modeled as individuated lexomes.

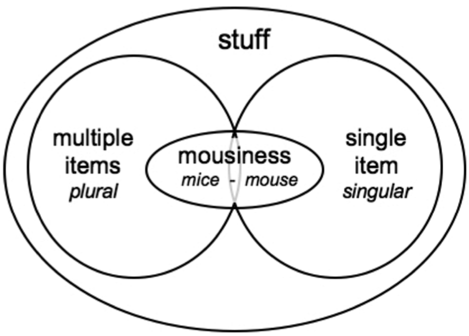

Figure 2: The relative specificity of the four dimensions as cues to plural forms. While the less specific cues (*stuff* and *mousiness*) will receive positive reinforcement early in learning, their ubiquity will cause them to produce more error than the uniquely informative cues. As a result, the influence of the less specific cues will wane as experience grows.

Figure 2 shows how the various potential semantic cues to "mice" overlap relative to a fairly simple set of lexomes – comprising irregular forms, regular stems, and a single lexome marking the regular plural contrast +*s* – in an idealized young learner. As can be seen, the structure of the lexicon is such that learning to correctly produce "mice" will require unlearning a number of irrelevant semantic cues.

As a child learns from this distribution of lexomes, the semantic dimension *multiple items* will initially be reinforced whenever plural forms are used to talk about any set of objects in context. In this distribution, most noun forms employ

Brought to you by | De Gruyter / TCS
Authenticated
Download Date | 7/3/18 7:37 PM

largely the same phonetic form to denote kind semantics (e.g., various articulations of "rat" are all strongly associated with rat semantics). It is thus the presence or absence of a terminal sibilant (+*s*) that discriminates between singularity and plurality when the noun stem is used. This means that initially, whenever a set of objects is to be described, the child's language model will predict that a form should have a final sibilant.

However, this expectation will change over time. Consider: whenever mouse objects are talked about, the relationship between *multiple mouse objects* and "mice" will be reinforced, as well as that between *multiple items* and "mice". This means that in other contexts where plurals are used, *multiple items* will not only lead the child to implicitly expect a final sibilant, it will also cause "mice" to be expected. When the child does not hear "mice", this will lead to error, downgrading the relationship between *multiple items* and "mice" in the child's language model. Similarly, whenever "mice" are talked about, the relationship between *multiple items* and a terminal sibilant will be weakened. Over time this will gradually increase the association between *multiple mouse objects* and "mice" in mice contexts, while decreasing the association between *multiple objects* and "mice" in non-mice contexts. Thus, the distribution of evidence in the linguistic environment will support discrimination of the appropriate cues to "mice".

So far we have discussed what the expectations of a naïve observational learner might be. A key question is how this will be influenced by production. Here we can envisage two scenarios: In the first, production serves to reinforce itself, such that when the noise in a child's underlying model of the world results in her saying "mouses," "mouses" is reinforced as a response. In the second case, production is driven by a child's model of her intended behavior (what she has observed), but it is the *attempts* to perform a behavior that reinforce the underlying model, not whatever noisy behavior emerges from them. Model-based learning, which we adopt here, can explain why practice leads ballroom dancers towards the right steps, rather than reinforcing treading on toes (Gläscher et al., 2010).

Behaviorally, of course, we know English speakers often do go through a period of saying "mouses" in childhood, but as they grow older, they come to produce only the adult form, "mice". In our model, we assume that this reflects an interim state that arises from the distributional properties of the learning environment, and is resolved by further sampling from this distribution. This is, of course, one of the vital advantages of formal models – they generate unambiguous predictions.

Ramscar and Yarlett (2007) present a computational simulation that reinforces a model of the world rather than the overt behavior it would generate, and which predicts that eliciting over-regularized forms from children will actually cause them to over-regularize *less*. In a series of experiments, Ramscar and Yarlett then show that children exhibit the very behavior predicted by the model: When seven-year-old children repeatedly produced the same plurals across blocks of trials, their rates of over-regularization went down in later blocks. This occurred even when children

Brought to you by | De Gruyter / TCS
Authenticated
Download Date | 7/3/18 7:37 PM

were given positive feedback on the incorrect forms they produced, lending further support to the idea that learning reinforces children's models of the world, rather than their behavior per se.

Ramscar et al (2013d) show that when the challenges facing a child language learner are explicitly set up in the way shown in Figures 1 and 2, the distribution of forms and semantics in English invariably leads to what has been described as "U-shaped" performance in plural production: Mastery of correct irregular forms is preceded by a phase in which both correct and incorrect irregular plurals are produced. Moreover, Ramscar et al show that in this model, the ultimate elimination of interference from sibilant final forms – which give rise to over-regularization – is driven by error caused by the inappropriate expectation of irregular forms when the semantics of *regular* forms are present in a lexical context. That is, the same non-discriminative semantic dimension that causes children to expect a sibilant final form in an irregular context – leading to over-regularization – causes them to erroneously expect irregular forms in regular contexts, gradually causing this non-discriminative dimension a cue to be unlearned as cues to irregulars, thereby reducing over-regularization.

This results in an unambiguous prediction: Engaging children at an appropriate stage of development in a task invoking the semantics of regular forms ought to bring about a reduction in over-regularization. To test this, Ramscar et al first pre-tested children on a task that elicited both regular and irregular plural and singular forms. One group of children then performed a color-related control task, while the other performed a memory task involving the same regular plural forms from the elicitation task along with regular lures. The children were then post tested using the elicitation task.

Predictably, the color task had no effect on over-regularization. However, consistent with the detailed predictions of the model, the younger children tested in the memory condition showed a small but significant *increase* in over-regularization. By contrast, the same exposure to the semantics of regular plural forms brought about a large and significant *decrease* in over-regularization in the older children (Ramscar et al. 2013a), just as the model predicts. In other words, the pattern of children's over-regularization and their retreat from it is exactly as one would predict given the way that children learn, the forms that they learn, and the way that these forms are distributed semantically.

5 The scientific appeal of discrimination learning

Because children reliably go through a period in which they over-regularize, and because they reliably stop over-regularizing without encountering any explicit instruction to do so, it has often been argued that their behavior presents a logical puzzle: Why would they ever stop (Pinker 1984, 2004; Pinker and Prince 1988)?

Brought to you by | De Gruyter / TCS
Authenticated
Download Date | 7/3/18 7:37 PM

The models and the results reported here show that the patterns of over-regularization behavior exhibited by developing English speakers are not puzzling. On the contrary, over-regularization is a direct byproduct of the distributional properties of the English language, and in typically developing children, the same processes and circumstances that give rise to the problem ultimately resolve it. Thus, over-regularization is a self-correcting problem, with little need for explicit correction. (Indeed, because learning reinforces children's models of the world, explicit feedback about behavior often has little obvious effect on what children actually learn; Ramscar and Yarlett, 2007.)

In the light of the debate that has surrounded this phenomenon, it seems worth noting that our explanation does not "solve" the puzzle of over-regularization as it has previously been posed. There are two reasons for this: First, because, while this puzzle has frequently been framed as a "logical problem", it is only actually a problem if one ignores some well-established facts about animal learning (Rescorla and Wagner, 1972), or if one assumes that human infants are somehow less capable learners than the animals these principles were derived from – namely, rats (Pinker, 2004).

And second, because once one understands how learning works, it becomes clear that the way the problem has been posed in the past makes little sense. Learning is not a process that "associates" "units" of forms and meanings in the way that morphologists have traditionally imagined (Ramscar and Port, 2015), but is instead a discriminative process in which *systems* of associations are learned implicitly, as a result of the process of dissociating anything and everything else. Learning has evolved to enable humans and other animals to make sense of the world by reducing its dimensionality and complexity in order to highlight what is relevant (Trimmer et al., 2012).

While it is not always easy to intuit exactly how this process allows us to map our semantic model of the world onto the system of forms in a language, it is possible to model this complex process computationally, and thereby gain a richer understanding of it. By contrast, not only does it make little sense from a learning perspective to imagine that language learners are faced with the task of acquiring rules that transform or agglutinate "form units" that map to "meaning units" – as morphologists have traditionally imagined – but, as the numerous claims that have been put forward for this or that aspect of linguistic processing being innate over recent years would seem to attest, conceiving of how this kind of combinatoric system actually works in any kind of detail appears to be all but impossible.

6 Discrimination learning and morphology

It is often assumed that regularity is a desirable or normative goal for morphological systems, and that irregular paradigms represent deviations from the uniform patterns that systems (or their speakers) strive to maintain. Such an assumption is challenged,

Brought to you by | De Gruyter / TCS
Authenticated
Download Date | 7/3/18 7:37 PM

however, by phenomena like suppletion, in which an inflected stem-change produces a phonologically unrelated allomorph (e.g., "mouse" / "mice"), rendering patterns of form-meaning mappings unpredictable.

A discriminative perspective makes precisely the opposite assumption. In discriminative models, the difference between overtly suppletive forms (such as "mouse" / "mice") and more regular forms (such as "rat" / "rats") is that the former serve to accelerate the rate at which a speakers' representation of a specific form/meaning contrast becomes discriminated from the form classes that express similar contrasts. The logic of these models is that all learning serves to increase the level of suppletion in a system of form-meaning mappings.

From this perspective, suppletive irregular forms (like "mice") are not categorically different types, as morphologists have been wont to imagine, but are merely extreme instances of the system of discriminative contrasts that linguistic communication relies on. Moreover, when one examines language use from this perspective, it is clear that these systematic contrasts are ubiquitous at the sub-phonemic level. Thus for example, while the combinatorial paradigm has long assumed that the forms of language are constructed out of an alphabet of phones (Port and Leary, 2005), numerous studies have made clear how markedly this idealization departs from the empirical truth. For example, the duration and fundamental frequency of the *cap* in "captain" differs systematically from the morphologically unrelated "cap". Moreover, analyses have shown that the nature of linguistic contrasts is such that even so-called homophones – like "time" and "thyme" – appear to be gesturally and acoustically distinct (Gahl, 2008).

The patterns of contrast observed in the lexicon are also observable at a morphological level: Baayen et al. (2003) found that a sample of speakers produced Dutch nouns with a longer mean duration when they occurred as singulars than when they occurred as the stem of the corresponding plural. Kemps et al. (2005) show that speakers are sensitive to these prosodic differences, finding that "acoustic differences exist between uninflected and inflected forms and that listeners are sensitive to them" (Kemps et al. 2005: 441). Plag et al. (2014) observed similar contrasts and sensitivities to them in a study of phonemically identical affixes in English.

It thus follows that from a discriminative perspective, it is the *regularity* in morphological systems that stands in need of explanation. Discrimination learning suggests a solution here as well. Unlike derivational processes, inflectional processes are traditionally assumed to be highly productive, defining uniform paradigms within a given class. Lemma size is thus not expected to vary, except where forms are unavailable due to paradigm 'gaps' or 'defectiveness'. However, corpus studies suggest that this expectation is an idealization. Many potentially available inflected forms are unattested in corpora, and as corpus sizes increase, sampling does not converge on uniformly populated paradigms, but rather reinforces classes and develops longer tails (Blevins et al., 2015).

Brought to you by | De Gruyter / TCS
Authenticated
Download Date | 7/3/18 7:37 PM

In order for a collection of partial samples to allow for the generation of unattested forms, the forms that speakers do know must be organized into systematic structures that collectively enable the scope of possible variations to be realized. These structures thus correspond to lexical neigbourhoods, whose effects have been investigated in a wide range of psycholinguistic studies (Baayen et al. 2006; Gahl et al. 2011). From the present perspective, these neighbourhoods are not independent dimensions of lexical organization. Instead they constitute the creative engine of the morphological system, permitting the extrapolation of the full system from partial patterns. Regular paradigms thus enable language users to generate previously unencountered forms in the same way as is captured by our model of plural morphology. Regular forms are not the product of an explicit rule, or of any kind of explicit grammatical knowledge, but rather, they are *implicit* in the distribution of forms and semantics in the language as a system.

7 Discriminative language learning

Traditionally, linguistic and social learning have been painted as being in opposition to "mere associative learning" (e.g., Tomasello, 2003), and most scientists and practitioners likely still believe this to be the case. However, these contrasts ultimately rely on a faulty understanding of what associative learning actually is (Rescorla, 1988). As we have highlighted above, associative learning is in fact a *discriminative* process, and one which is inherently systematic. While the systematicity of learning is simple to state, the explanatory tools it provides are both subtle and powerful.

In illustrative work, Ramscar et al (2013c) have shown that changes across the lifespan in adults' ability to learn the association of arbitrary pairs of words such as *jury* and *eagle* – while at first glance an almost prima facie example of a combinatorial process – are far better predicted and modeled in terms of systematic discrimination learning in the lexicon. For instance, it is well attested empirically that while learning frequently co-occurring pairs (like *lock–door*) differs little with age, learning unlikely pairs like *jury-eagle* becomes increasingly difficult. From a discriminative perspective, the explanation for this is straightforward: The fact that the latter pair only co-occurs rarely causes them to become negatively associated in the lexical system as a whole, and these negative weights increase with experience. Thus, although learning to pair *jury* and *eagle* appears to be a combinatoric process, it turns out that actual behavior in the task is best explained by modeling the lexicon as a densely interconnected system in which <u>all</u> lexical items are related by complex patterns of co-occurrence, and in which learning causes items to become associated or dissociated from one another as a function of experience (i.e., in the same way as morphology was treated in our model above). When paired associate learning is modeled as a task that requires its subjects to reverse the systematic *dis*sociations that ordinary experience teaches (i.e., that *jury* is not informative about *eagle*), the

Brought to you by | De Gruyter / TCS
Authenticated
Download Date | 7/3/18 7:37 PM

changing ability of adults to learn these pairs at various ages can be predicted with surprising accuracy (Ramscar et al 2013c).

By highlighting the systematic nature of language learning, this view of morphological development sheds new and productive light on the many apparent puzzles that arise when the learning of aspects of linguistic systems is considered in isolation. It is because of this systems-level focus that this approach will ultimately yield fruitful methods for understanding how and when language learning goes awry: As Quine (1960) noted, learning language requires that a child master not only the relationship between a system of conventionalized sound and meanings relationships, but also that the child learn how to use this system to communicate. Being able to do so appears to hinge on learning to share subjectivity; the child must somehow learn to comprehend the shared point of view of her community (see also Tomasello, 2003; Wittgenstein, 1953).

Exactly how human infants come to discriminate the "intersubjectively available cues as to what to say and when" (Quine, 1960) is an incredibly complex task. A discriminative account of communication grounded in learning theory allows us to frame questions about the way that children learn the sets of conventionalized cues that underpin languages (Wittgenstein, 1953) in ways that are tractable and amenable to formal description. For example, if communicative conventions – and language – are the product of learning, why is language apparently solely the preserve of humans? What is *special* about human learners?

One part of the answer to this question lies in the difference between the way that the brains of humans and other animals develop, and its impact on learning. Like many other primates, humans are born with immature brains. Birth is followed by synaptogenesis (the proliferation of synapses) followed by an extended pruning period (synaptic elimination). Brain development in humans, however, is markedly different from that of other primates. In monkeys, the postnatal development of the brain occurs at the same rate in all cortical areas. In contrast, human cortical development is uneven: Synaptogenesis in the visual and auditory cortex peaks a few months after birth, while the same developments occur later in the prefrontal cortex, which doesn't fully mature until late adolescence (Ramscar and Gitcho, 2007; Thompson-Schill, Ramscar, and Evangelia, 2009).

One important behavioral consequence of delayed prefrontal development is young children's inability to select behaviors that conflict with prepotent responses (Ramscar, et al. 2013b). In adults, prefrontal control mechanisms bias responses and attention according to goals or context, selectively maintaining task-relevant information and discarding task-irrelevant information (Ramscar and Gitcho, 2007; Thompson-Schill, Ramscar, and Evangelia, 2009). The absence of this capacity in young children can be illustrated by contrasting their performance with that of adults on biased selection tasks, such as guessing the hand an M&M is in. When the hands are biased 25:75, children up to age 5 tend to overmatch, fixating on the high-probability ''good'' hand. After age 5, however, a probability matching strategy emerges (Derks

Brought to you by | De Gruyter / TCS
Authenticated
Download Date | 7/3/18 7:37 PM

and Paclisanu, 1967). This is a rare instance in which children's inability to think flexibly is an advantage – probability matching actually reduces the number of M&Ms won.

Another area of learning in which cognitive flexibility may well prove disadvantageous is in the process of learning even a "simple" morphological system like that of English noun inflection. Linguistic knowledge is, in its essence, conventional. In the presence of a linguistic cue, a social animal needs to be able to understand or respond appropriately given the context. For this to happen, linguistic signals, must be both conventionalized and internalized (Wittgenstein, 1953). Learning the system of cues that yields appropriate understanding is far more likely to happen if learners are unable to filter their attention during the course of learning. Given a similar set of cues and labels to learn, young learners will tend to sample the environment in much the same way, and thus are more likely to develop similar expectations regarding the relationship between cues and symbols.

In contrast to children, adults struggle to master linguistic conventions (including English noun morphology; Johnson and Newport, 1989). This may reflect an inevitable handicap brought about by their increased ability to selectively attend and respond to the world and the cues in it. Because development increases the complexity of the human learning architecture, allowing learners to filter their attention in learning, it is likely that it also dramatically reduces the ability to learn conventions by naively sampling in the manner described above (Ramscar and Gitcho, 2007; Thompson-Schill et al., 2009). Put simply, the greater variety there is in what adults attend to during learning, the less conventionality there will be in what adults learn. Conversely, the less that children are able to direct their attention in learning, the more what they learn will be shaped by their immediate physical, social, and linguistic environments, and the more their learning about common regularities in these environments will be conventionalized (see also Finn et al., 2013, 2014; Hudson Kam and Newport, 2005, 2009; Singleton and Newport, 2004).

Although discriminative rules can be used to model some of the brain's learning processes, it is clear that there is far more to learning than error monitoring. Humans, especially adult humans, are not the passive observers of the environment that Rescorla-Wagner idealizes them to be. Our understanding of exactly how attention and learning trade off against one another, and how this affects what gets learned as frontal regions mature is, as yet, in its infancy. However, the learning and control processes we describe are amenable to computational and biological modeling, and progress is being made in this regard (Ramscar et al., 2013c).

To emphasize the point that an idealized model can be useful even when incomplete, consider that Triesch, Teuscher, Deák, and Carlson (2006) have shown how gaze following emerges naturally from discriminative learning mechanisms, provided that a child has access to a caregiver that tends to look at things in a ways the infant finds informative. A model like this is far more likely to lead to an understanding of how social development goes awry, and how it impacts on other

Brought to you by | De Gruyter / TCS
Authenticated
Download Date | 7/3/18 7:37 PM

aspects of learning, than the many current theories that propose that gaze following is the result of an unspecified innate mechanism. As Box and Draper (1987) famously noted, while models are always wrong in the limit, they can still be very useful.

8 Conclusion

While the studies reviewed here have uniformly concerned themselves with normal language development, the discriminative approach has the potential to provide a raft of new insights into the impaired development that is characteristic of language disorders. Normal developmental trajectories depend both on stable learning processes, and on broad sampling of the linguistic environment. When sampling is impoverished, either because of the surrounding linguistic environment or because of perceptual deficits, such as hearing loss, trajectories may be predictably slowed or impaired. Learning may also be disrupted if the coordinative processes that subsume this type of sampling are in some way compromised (e.g., as a result of idiosyncrasies in attentional mechanisms, in the timing of prefrontal development, or in the social dynamics of parent-child interaction; Gros-Louis, West, and King, 2014; Warlaumont et al. 2014).

By assessing the structure of the linguistic input and the dynamics of the learning process, discriminative models furnish the analytic tools to better investigate and isolate the myriad causes of disordered development. Such analyses can pinpoint where the child is on the learning trajectory (e.g., identifying the difference between younger and older learners based on their errors; Ramscar et al. 2013d) and also suggest possible interventions to help move them along (such as adopting postnominal constructions to facilitate color and number learning, Ramscar et al. 2010; 2011; or highlighting optimal timing dynamics between parent and child in labeling, Yu and Smith, 2012). Many other such other targeted interventions are possible; the space of possible applications has only begun to be explored.

We have sought here to provide an overview of what a discriminative learning approach to language learning looks like, along with the benefits that adopting formal and conceptual models of discriminative learning can bring to our understanding of human development. It is our hope that more of our colleagues will be inspired to consider this approach in the future.

References

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

Brought to you by | De Gruyter / TCS
Authenticated
Download Date | 7/3/18 7:37 PM

Baayen, R. Harald, Laurie Beth Feldman & Robert Schreuder. 2006. Morphological influences on the recognition of monosyllabic monomorphemic words. *Journal of Memory and Language* 55(2). 290–313.

Baayen, R. Harald, James M. McQueen, Ton Dijkstra & Robert Schreuder. 2003. Frequency effects in regular inflectional morphology: Revisiting Dutch plurals. In Baayen, R. Harald & Robert Schreuder. (eds.), *Morphological structure in language processing*, 355–370. Berlin: Mouton de Gruyter.

Blevins, Jim, Petar Milin, & Michael Ramscar. 2015. Zipfian discrimination. *NetWordS*.

Box, George E.P. & Norman Richard Draper. 1987. *Empirical model-building and response surfaces*. New York: Wiley.

Davies, Mark. 2009. The 385+ million word corpus of contemporary American English (1990-present). *International Journal of Corpus Linguistics* 14. 159–90.

Davis, Matthew H., William D. Marslen-Wilson & M. Gareth Gaskell. 2002. Leading up the lexical garden path: Segmentation and ambiguity in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance* 28(1). 218.

Derks, Peter L. & Marianne I. Paclisanu. 1967. Simple strategies in binary prediction by children and adults. *Journal of Experimental Psychology* 73(2). 278–285.

Finn, Amy S., Taraz Lee, Allison Kraus & Carla L. Hudson Kam. 2014. When it hurts (and helps) to try: The role of effort in language learning. *PloS one* 9(7). e101806.

Finn, Amy S., Carla L. Hudson Kam, Marc Ettlinger, Jason Vytlacil & Mark D'Esposito. 2013. Learning language with the wrong neural scaffolding: the cost of neural commitment to sounds. *Frontiers in systems neuroscience*, 7.

Gahl, Susanne. 2008. Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language* 84(3). 474–496.

Gahl, Susanne, Yao Yao & Keith Johnson. 2012. Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language* 66(4). 789–806.

Gläscher, Jan, Nathaniel Daw, Peter Dayan & John P. O'Doherty. 2010. States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron* 66(4). 585–595.

Gros-Louis, Julie, Meredith J. West & Andrew P. King. 2014. Maternal responsiveness and the development of directed vocalizing in social interactions. *Infancy* 19(4). 385–408.

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

Kam, Carla L. Hudson & Elissa L. Newport. 2009. Getting it right by getting it wrong: When learners change languages. *Cognitive Psychology* 59(1). 30–66.

James, William. 1890. *The principles of psychology*. New York: Henry Holt.

Johnson, Jacqueline S. & Elissa L. Newport. 1989. Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology* 21. 60–99.

Kamin, Leon J. 1969. Predictability, surprise, attention, and conditioning. *Punishment and Aversive Behaviour*. 279–296.

Kemps, Rachèl JJK, Mirjam Ernestus, Robert Schreuder & R. Harald Baayen. 2005. Prosodic cues for morphological complexity: The case of Dutch plural nouns. *Memory & Cognition* 33(3). 430–446.

Montague, P. Read, Steven E. Hyman & Jonathan D. Cohen. 2004. Computational roles for dopamine in behavioural control. *Nature* 431(7010). 760–767.

Miller, Ralph R., Robert C. Barnet & Nicholas J. Grahame. 1995. Assessment of the Rescorla-Wagner model. *Psychological bulletin* 117(3). 363.

Brought to you by | De Gruyter / TCS
Authenticated
Download Date | 7/3/18 7:37 PM

Niv, Yael. 2009. Reinforcement learning in the brain. *Journal of Mathematical Psychology* 53. 139–154.
Pavlov, Ivan. 1927. *Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex*, trs. Gelb Anrep. London: Oxford University Press.
Pinker, Steven. 2004. Clarifying the logical problem of language acquisition. *Journal of Child Language* 31. 949–953.
Port, Robert F. & Adam P. Leary. 2005. Against formal phonology. *Language* 81(4). 927–964.
Plag, Ingo, Julia Homann & Gero Kunter. 2015. Homophony and morphology: The acoustics of word-final S in English. *Journal of Linguistics* 53(1). 181–216.
Quine, W. V. O. 1960. *Word and object*. Cambridge, MA: MIT Press
Reddy, Michael. 1979. The conduit metaphor. *Metaphor and thought* 2. 164–201.
Ramscar, Michael. 2013. Suffixing, prefixing, and the functional order of regularities in meaningful strings. *Psihologija* 46(4). 377–396.
Ramscar Michael & R. Harald Baayen. 2013. Production, comprehension, and synthesis: a communicative perspective on language. *Frontiers in Psychology* 4. 233. doi:10.3389/fpsyg.2013.00233.
Ramscar, Michael & Melody Dye. 2011. Learning language from the input: Why innate constraints can't explain noun compounding. *Cognitive Psychology* 62(1). 1–40.
Ramscar, Michael, Melody Dye, Hanna Muenke Popick & Fiona O'Donnell-McCarthy. 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(7). e22501. doi:10.1371/journal.pone.0022501
Ramscar, Michael, Melody Dye, Jessica W. Gustafson & Joseph Klein. 2013. Dual routes to cognitive flexibility: Learning and response conflict resolution in the dimensional change card sort task. *Child Development* 84(4). 1308–23.
Ramscar, Michael, Melody Dye & Joseph Klein. 2013. Children value informativity over logic in word learning. *Psychological Science* 24(6). 1017–1023.
Ramscar, Michael, Melody Dye & Stewart M. McCauley. 2013. Error and expectation in language learning: The curious absence of 'mouses' in adult speech. *Language* 89(4). 670–793.
Ramscar, Michael & Nicole Gitcho. 2007. Developmental change and the nature of learning in childhood. *Trends in cognitive sciences* 11(7). 274–279.
Ramscar, Michael, Peter Hendrix, Bradley Love & R. Harald Baayen. 2013. Learning is not decline: The mental lexicon as a window into cognition across the lifespan. *The Mental Lexicon* 8(3). 450–481.
Ramscar, Michael, Peter Hendrix, Cyrus Shaoul, Petar Milin & R. Harald Baayen. 2014. The myth of cognitive decline: Non-linear dynamics of lifelong learning. *Topics in Cognitive Science* 6. 5–42.
Ramscar, Michael & Robert Port. 2015. Categorization (without categories). In Dawbroska, Ewa & Dagmar Divjak (eds.), *Handbook of cognitive linguistics*, 75–99. Berlin, Boston: De Gruyter Mouton.
Ramscar, Michael & Daniel Yarlett. 2007. Linguistic self-correction in the absence of feedback: A new approach to the logical problem of language acquisition. *Cognitive Science* 31(6). 927–960.
Ramscar, Michael, Daniel Yarlett, Melody Dye, Katie Denny & Kirsten Thorpe. 2010. The effects of feature-label-order and their implications for symbolic learning. *Cognitive Science* 34(6). 909–957.
Rescorla, Robert A. 1968. Probability of shock in the presence and absence of CS in fear conditioning. *Journal of Comparative & Physiological Psychology* 66. 1–5.
Rescorla, Robert A. 1988. Pavlovian conditioning: It's not what you think it is. *American Psychologist* 43(3). 151–160.
Rescorla, Robert A. & Allan R. Wagner. 1972. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In Abraham H. Black & William F. Prokasy (eds.), *Classical conditioning II: Current research and theory*, 64–99. New York: Crofts.

Brought to you by | De Gruyter / TCS
Authenticated
Download Date | 7/3/18 7:37 PM

Rose, Tony, Mark Stevenson & Miles Whitehead. 2002. The Reuters Corpus Volume 1-From yesterday's news to tomorrow's language resources. *LREC* 2. 827–832.
Rumelhart, David E. & James L. McClelland. 1986. *On learning the past tenses of English verbs*. California University, San Diego: La Jolla Institute for Cognitive Science.
Saffran, Jenny R., Richard N. Aslin & Elissa L. Newport. 1996. Statistical learning by 8-month-old infants. *Science* 274(5294). 1926–1928.
Schultz, Wolfram. 2006. Behavioral theories and the neurophysiology of reward. *Annual Review of Psychology* 57. 87–115.
Schultz, Wolfram. 2010. Dopamine signals for reward value and risk: basic and recent data. *Behavioral and Brain Functions* 6. 1–9.
Seidenberg, Mark S. & David C. Plaut. 2014. Quasiregularity and its discontents: The legacy of the past tense debate. *Cognitive Science* 38(6). 1190–228.
Seidenberg, Mark S. & James L. McClelland. 1989. A distributed, developmental model of word recognition and naming. *Psychological review* 96(4). 523.
Shannon, Claude Elwood. 1948. A mathematical theory of communication. *Bell Systems Technical Journal* 27(3). 379–423.
Siegel, Shepard & Lorraine G. Allan. 1996. The widespread influence of the Rescorla-Wagner model. *Psychonomic Bulletin & Review* 3(3). 314–321.
Singleton, Jenny L. & Elissa L. Newport. 2004. When learners surpass their models: The acquisition of American Sign Language from inconsistent input. *Cognitive Psychology* 49. 370–407.
Thompson-Schill, Sharon L., Michael Ramscar & Evangelia G. Chrysikou. 2009. Cognition without control when a little frontal lobe goes a long way. *Current Directions in Psychological Science* 18(5). 259–263.
Todes, Daniel P. 2014. *Ivan Pavlov: A Russian life in science*. Oxford: Oxford University Press.
Tomasello, Michael. 2003. *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
Tomasello, Michael. 2008. *Origins of human communication*. Cambridge, MA: MIT Press.
Triesch, Jochen, Christof Teuscher, Gedeon O. Deák & Eric Carlson. 2006. Gaze following: Why (not) learn it? *Developmental Science* 9(2). 125–147.
Trimmer, Pete C., John M. McNamara, Alasdair I. Houston & James A.R. Marshall. 2012. Does natural selection favour the Rescorla–Wagner rule? *Journal of Theoretical Biology* 302. 39–52.
Waelti, Pascale, Anthony Dickinson & Wolfram Schultz. 2001. Dopamine responses comply with basic assumptions of formal learning theory. *Nature* 412. 43–48.
Warlaumont, Anne S., Jeffrey A. Richards, Jill Gilkerson & D. Kimbrough Oller. 2014. A social feedback loop for speech development and its reduction in autism. *Psychological Science* 25. 1314–1324.
Wittgenstein, Ludwig. 1953. *Philosophical investigations*. London: Blackwell.
Yu, Chen & Linda B. Smith. 2012. Embodied attention and word learning by toddlers. *Cognition* 125(2). 244–262.

Brought to you by | De Gruyter / TCS
Authenticated
Download Date | 7/3/18 7:37 PM