PSIHOLOGIJA, 2013, Vol. 46 (4), 377–396
UDC 159.9.07:81'373.611 ; 81'23 ; 519.17

DOI: 10.2298/PSI1304377R

Suffixing, prefixing, and the functional order of regularities in meaningful strings

Michael Ramscar

Department of Linguistics, University of Tübingen, Germany

The world's languages tend to exhibit a suffixing preference, adding inflections to the ends of words, rather than the beginning of them. Previous works has suggested that this apparently universal preference arises out of the constraints imposed by general purpose learning mechanisms in the brain, and specifically, the kinds of information structures that facilitate discrimination learning (St Clair, Monaghan, & Ramscar, 2009). Here I show that learning theory predicts that prefixes and suffixes will tend to promote different kinds of learning: prefixes will facilitate the learning of the probabilities that any following elements in a sequence will follow a label, whereas suffixing will promote the abstraction of common dimensions from a set of preceding elements. The results of the artificial language learning experiment support this analysis: When words are learned with consistent prefixes, participants learned the relationship between the prefixes and the noun labels, and the relationship between the noun labels and the objects associated with them, better than when words were learned with consistent suffixes. When words were learned with consistent suffixes, participants treated similarly suffixed nouns as being more similar than nouns learned with consistent prefixes. It appears that while prefixes tend to make items more predictable and to make veridical discriminations easier, suffixes tended to make items cohere more, increasing the similarities between them.

Key words: *discrimination learning, sequence learning, communication, language, morphology*

Across the world's languages, regularities that serve as modifiers to the specific properties of words tend to attach to more to their ends (as suffixes) rather than attaching to their beginnings (as prefixes) or being inserted into their middles (as infixes; see e.g., Sapir, 1921; St Clair, Monaghan, & Ramscar, 2009). In a survey, Hawkins and Gilligan (1988) found although the most common pattern for this kind of marking is a combination of prefixing and suffixing, languages that make exclusive use of suffixes (74 out of 203 languages studied) are far more common than languages that use only prefixes (9 of 203). Further, even among the languages that employ a combination of prefixes and suffixes, the latter tend to outnumber the former: (Fudge, 1984).

Various suggestions have been made the underlying causes of these patterns of prefixing and suffixing: Cutler, Hawkins, and Gilligan (1985; also

Corresponding author: ramscar@gmail.com

Hawkins & Cutler, 1988) suggest that there are offers processing benefits to be gained form identifying words in speech as soon as is possible, and given that prefixes provide little information about the unique identity of words, and actually serve to delay the arrival of uniqueness point information (Marslen-Wilson, 1987; Balling & Baayen, 2012), they are less communicatively efficient than suffixes, which do not delay the identification of word roots.

Another suggestion is that suffixes facilitate language learning as well as language processing, in particular with regards of marking the grammatical category of root words (Hawkins & Gilligan, 1988; St Clair et al., 2009). In an artificial grammar learning (AGL) task, St Clair et al found that markers learned as suffixes led to better learning of the relationship between marker words and category words than markers learned as prefixes (see also Hupp, Sloutsky, & Culicover, 2009).

St Clair et al, (pace Greenberg, 1957) suggested that suffixing and prefixing can be equated to "convergent" and "divergent" learning hierarchies, respectively (Osgood, 1949). In the framework of stimulus-response (S-R) associative learning, convergent hierarchies describe a variety of stimuli associated with a functionally identical response (S1, S2, ...Sx $\Rightarrow$ R), while divergent hierarchies associate similar stimuli with varied responses (S1 $\Rightarrow$ R1, R2, ...Rx). St Clair et al suggest that the former of these hierarchies is similar to suffixing, and the latter is similar to prefixing, and since (Osgood, 1949) noted that convergent hierarchies result in greater facilitation and positive transfer in learning the relationship between the stimuli and the response, whereas divergent hierarchies yield negative transfer and interference in learning (Osgood, 1949). St Clair et al suggest that, as a result,

> *"*informative associations between root words and suffixes will be more readily learned than those between prefixes and root words. Therefore, if natural languages are adapted for learning grammatical categories in terms of suffixing, then suffixes in natural languages ought to provide more reliable information about the category of the root word than prefixes, which provide information less easily available for learning*."* (St Clair et al., 2009, p 1319)

In what follows, I examine a slightly different way of conceiving of the learnability and informativity of suffixes and prefixes, based on the more detailed understanding of associative learning that has arisen in the half-century since Osgood's discussion of "convergent" and "divergent" hierarchies. I will suggest that the view of learning that has emerged from the associative tradition can not only inform our understanding of suffixing and prefixing, but also that it can help clarify our understanding of human communication itself.

I begin by describing contemporary learning theories in more detail, and lay out the difference between modern views of discrimination learning and associationism. I then use these theories to develop an analysis of suffix– and prefix– learning based on two basic principles derived from learning theory:

First, that learning is driven by uncertainty; and second, that as a result of the way that learning reduces uncertainty, learning is competitive, such that previous learning can block subsequent learning. I show that these basic principles suggest that suffix- and prefix- learning will yield qualitatively different information structures, such that linguistically suffix-learning should not be seen as being "better" than prefix-learning, but rather that the two should be seen as playing different – albeit rather complementary – functional roles. This analysis is used to derive predictions that are then tested, yielding encouraging empirical support for this view.

ASSOCIATIVE LEARNING IS NOT WHAT YOU THINK IT IS

Our understanding of associative learning has its origins in Ivan Pavlov's (Pavlov & Anrep, 1927) classical conditioning experiments. However, from a contemporary standpoint, it is important to note that the way associative learning is typically conceived of in the Psychology and Linguistics literatures is not just inconsistent with the view of animal learning that emerged from studies following up on Pavlov's initial discoveries, it is almost in complete opposition to it (Rescorla, 1988). Pavlov famously discovered that if a bell rang as food was presented to dogs, the dogs would soon began to salivate on hearing the bell when no food was on offer. This gave rise to a view of learning, based on *association* (we might call this "classical associative theory"). On this view, learning "associates" unrelated things in the world, such as a bell and a meal, by simply noting the degree to which a *stimulus* (e.g., the bell) and a *response* (e.g., salivating) are paired.

This naïve view of Pavlovian learning (a simple process that "computes nothing more than correlations," Santos, Flombaum, & Phillips, 2007) has been shown to be incapable of accounting for the actual facts of animal learning (Rescorla, 1988), as have two popular – but equally false – beliefs about what is necessary and sufficient for learning:

1. That explicit "rewards" or "punishments" are *necessary* for learning.
2. That co-occurrences between a "stimulus" and a "response" are *sufficient* for learning.

Studies have clearly demonstrated both of these beliefs to be false, and have shown that classical associative theory cannot explain the learning that occurs in animal conditioning.

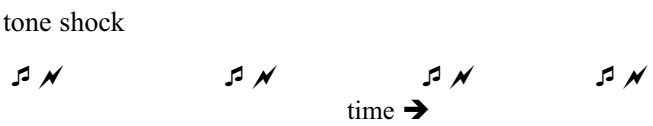

Figure 1. Schematic of a conditioning schedule used in Rescorla (1968). The rate of tones absent shocks here is 0.

For example, Rescorla (1968) trained a group of rats using the schedule depicted in Figure 1. These rats learned to associate a tone with a shock (and froze on hearing later tones), however other rats exposed to an *identical* number of tone-shock pairings, but into which tones that were not followed by shocks were interpolated (Figure 2) showed a different pattern of learning: Conditioning on these latter rats decreased as the rate of the tones without shocks increased.

Figure 2. A training schedule in which the background rate of tones absent-shocks increases. Although the absolute number of tones leading to shocks is identical, only 1 in 3 of the tones is now followed by a shock, and the degree to which rats condition to the relationship between the tones and the shocks diminishes proportionally (Rescorla, 1968).

Given that there was no change in the association rate between the groups of rats – only the background rate varied – it follows that this finding cannot be explained the naïve, classical associative theory described above (see also Rescorla, 1988; Ramscar, Dye, & Klein, 2013). Rather, the differences in learning observed must have been due to the "no shock" trials. That is, the *non*-occurrence of (expected) shocks after no-shock tones must influenced the degree to which the rats conditioned to the tones that did precede shocks.

It follows from this result that learning cannot simply be a process of tracking positive co-occurrences between cues and events. Similarly, the actual learning process has been shown to be more that a process simply counting successful and unsuccessful predictions. Associative learning in animals is a process that serves to reduce *uncertainty* about the predictive structure of an animal's environment (Rescorla, 1988). Because uncertainty is finite (it reduces as cues are learned and reliable expectations are formed), learning has been revealed to be a competitive process: As an animal learns to predict an outcome from a set of cues, this reduces that amount of uncertainty available to drive the learning of other cues. *Cue competition* is a simple statistical consequence of this process, and is illustrated in the results of *blocking* experiments (Kamin, 1969), which have shown that learning about the predictive value of a novel cue can be effectively 'blocked' by the presence of an already well-learned cue.

For example, a rat that has already learned that it will be shocked when it hears a tone will fail to learn to value a light that is introduced into its training schedule alongside the tone as an additional predictive cue. This is because the tone is already fully informative about the upcoming shock, and so the information provided by the light is redundant and ignored by the learning process (prior learning about the tone is said to "block" learning about the light).

Numerous results like the ones described here have shown that animals do not learn to "associate" stimuli and responses in the way many scientists still naively suppose (Rescorla, 1988). Rather, animals learn to discriminate the degree to which cues are informative about the environment. Cue competition systematically uncovers any positively informative relationships within an animal's

environment by eliminating the influence of *less* informative relationships. And since, invariably, the number of uninformative coincidences in the environment will far outnumber the informative ones, it follows that expectations that are *wrong* have more influence on the shape of this discriminative learning process than expectations that are *right* (this is why this process is often referred to as "error-driven learning").

"Associative learning" has thus come to be understood computationally as a discriminative process in which – in principle – everything in an animal's local environment potentially matters in predicting upcoming events (Ramscar et al., 2010). Although for the sake of simplicity models and explanations usually focus on informative cues and ignore cues whose high background rates are likely to render them largely irrelevant in competitive terms, in reality, prior learning influences – and is integral to – subsequent learning as part of an embodied, dynamic system in which what an animal learns in a given context must be understood against the backdrop of what it has already learned (Ramscar et al., 2010; Rescorla, 1988; this also helps clarify why learning is often related to a "stimulus complex," rather than individual stimuli; Rescorla & Wagner, 1972).

Finally, the logic of discrimination learning in turn suggests that at the outset, a learner's "knowledge" can be seen as comprising a large, undifferentiated set of cues associated with few or no environmental features (for modeling purposes, one might initially idealize this as comprising no more than "the world," i.e., the initial set of outcomes = 1), and that perceptible variances in the environment, along with a learner's developing expectations regarding them, drive the discrimination of the combination of cue values that best predict environmental features and their saliency (Rescorla, 1988). While this view is very different to the way learning is normally conceived of in contemporary Linguistics and Psychology, it is remarkably similar to William James' (1890) description of the way an infant first experiences the world, and of the way that the perception of variance leads her to learn to discriminate its contents:

> "the undeniable fact being that *any number of impressions, from any number of sensory sources, falling simultaneously on a mind* ***which has not yet experienced them separately****, will fuse into a single undivided object for that mind*. The law is that all things fuse that *can* fuse, and nothing separates except what must... Although they separate easier if they come in through distinct nerves, yet distinct nerves are not an unconditional ground of their discrimination, as we shall presently see. The baby, assailed by eyes, ears, nose, skin, and entrails at once, feels it all as one great blooming, buzzing confusion; and to the very end of life, our location of all things in one space is due to the fact that the original extents or bignesses of all the sensations which came to our notice at once, coalesced together into one and the same space. There is no other reason than this why "the hand I touch and see coincides spatially with the hand I immediately feel" James (1890, p488; emphases in original).

Like associative learning, James' "blooming, buzzing confusion" is frequently mischaracterized in the literature; however, discriminative learning from error offers the best account of the process through which the perception of the variance described by James drives animals' learning about the world. Moreover, the computational properties of this process has been extensively explored (McLaren & Mackintosh, 2000; Dickinson, 1980; Pearce & Hall, 1980; Rescorla & Wagner, 1972; see Danks, 2003 for a review), and considerable progress has been made in understanding its biological underpinnings (Montague, Hyman, & Cohen, 2004; Niv, 2009; Schultz, 2006; Schultz, 1998; Schultz, Dayan, & Montague, 1997; Schultz & Dickinson, 2000; Waelti, Dickinson, & Schultz, 2001).

SUFFIXES, PREFIXES AND INFORMATION STRUCTURE IN LEARNING

In a series of studies (Ramscar, Yarlett, Dye, Denny, & Thorpe, 2010, Ramscar, Dye, Muenke Popick, & O'Donnell-McCarthy, 2011; Ramscar, Dye, Gustafson, & Klein, 2013) my colleagues and I have shown how the sequence in which words appear can exert a powerful influence on what a child learns about them. Consider, for example, a child learning about color. Although color words appear in children's vocabularies from a very young age, sighted children's early use of them is comparable to that of blind children: that is, they can produce them in familiar contexts ("yellow banana"), but cannot pick out novel objects by color, or reliably apply color words in unfamiliar contexts (Ramscar et al., 2010). It appears that children struggle to grasp the way that *specific words* match to *specific hues*. Why?

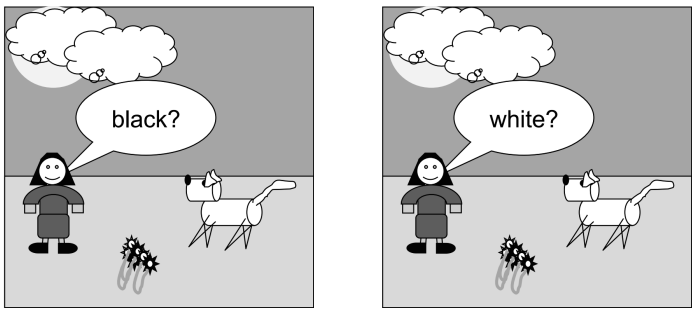

Figure 3. The ubiquity of color poses a number of problems for a child trying to learn color words. How is a child to learn what "black" and "white" mean in situations like this?

The first problem that a child learning color words *must* overcome is that she will never encounter color independently: she may encounter green apples, or green grass, but she will never encounter green on its own (Wittgenstein, 1953). To further complicate matters, it is virtually impossible to ascertain the meaning of a given color word from a single encounter. For example, for a child faced with the scene shown in Figure 3, the cues to the words "black" and

"white" and "gray" etc. will initially be identical. This creates a discrimination problem: over time, a child must learn to discriminate which features of the world appropriately match a given word in a given context.

The discriminative view of learning described above offers a way of explaining how this might happen. Color word learning requires that color (or hue) should be treated as the most reliable cue to a given color word, discriminating it from other less reliable competitors (such as alternate hues and other object features). Although this task might appear impossible when color words are heard in isolation (Figure, 3), if words are heard sequentially (as they are), then once a child knows what a dog is and is able to associate it with the label "white," she can learn to discriminate the cues that predict a label – such as "white" – via competitive learning amongst features (Figure, 4).

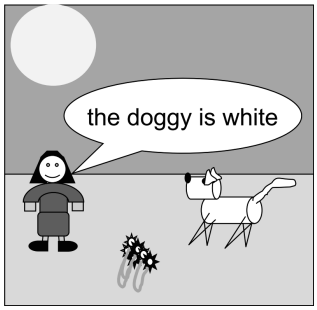

Figure 4. Once a child knows the names of things like dogs, language can help focus her attention to the colors of just those things, and this will make the task of learning the meanings of color words easier (note that this still leaves the child with the problem of figuring out what about the doggy is "white")

If the child has heard "Look! That doggy is white," in the presence of a white dog, then the next time she sees a dog, and hears like, "Look! That doggy is..." she will be likely to expect to hear "white". (NB. this expectation is implicit: she won't be consciously be thinking, "Oh, hello, now here comes *white*").

If this dog is also white, and she hears "white", then while this will help her to strengthen the connection between dog and "white", it won't help her learn what "white" means. To do this, she needs *error.* Suppose the next dog is brown: Now, when she hears, "Look at the doggy! He's..." she will be expecting to hear "white." Because the expectation that she will hear "white" is erroneous (she hears "brown"), she will learn to devalue all the *other* features of the dog that she had erroneously supposed were cues to "white" (the wet nose, waggy tail, fur, etc.); i.e., she will learn that she is less likely to hear "white" when these things are present than she had supposed. This in turn will cause value to shift from features that produce more error to those that produce less: white will be implicitly strengthened as a cue to "white" simply because all of the other dog features have been devalued as cues to the label "white." Despite the fact that she neither heard the word "white" nor saw any white, the child's understanding of the relationship between *white* and "white" will have improved.

While color word learning works well when color features predict a discrete Label (**FL**-learning; Ramscar et al., 2010; in Osgood's, 1949, terms, this situation can be understood as the features converging to predict the label), if this sequence is temporally reversed (i.e., "Look at the white doggy!"), such that the process becomes one of learning to predict a complex set of Features from a discrete Label (**LF**-learning; i.e., in Osgood's terms, the learning hierarchy is divergent), then things change. Importantly, competition between cues *cannot* occur in this situation, since, linguistically, the label is the only cue present (value cannot transfer to other cues when there are no other cues; Ramscar et al., 2010). Although **FL**- and **LF**-learning appear similar, the difference in their temporal sequencing results in markedly different information structures, and very different patterns of learning.

In **FL**-learning, features compete, and unreliable features lose value to the most reliable feature. By contrast, in **LF**-learning, competitive learning amongst features is not possible. There is, in effect, just a single cue is present – the label – and the label's features will not covary in any meaningful way with the outcomes (all of the features of "white" will tend to be present whenever "white" is uttered). As a consequence, **FL**-learning will tend to promote a simple, *probabilistic* representation of the relationship between a label and the features of object labels it predicts (specifically, the co-occurrence probability between the label and each feature, normalized by the probability of the label), but in the absence of cue competition, discrimination will be poor. Consistent with this analysis, Ramscar et al. (2010) found that training color words with with postnominal constructions (FL) significantly improved the accuracy and consistency of two-year olds' color word application, whereas a similar schedule of prenominal training (LF) had no effect on performance at all (see Ramscar et al, 2013 for a replication, and Ramscar et al., 2011 for similar results in a number learning task).

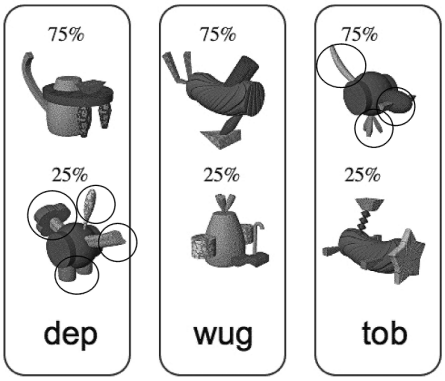

Figure 5. Examples of the category stimuli and structures used in Ramscar et al. (2010), Experiment 1. Notably, the 'high-saliency' body type does not help distinguish between categories. The sets of discriminating features that need to be learned in order to successfully distinguish the subcategories are circled on the low-frequency ''dep'' and high-frequency ''tob'' exemplars.

Similar results were obtained in a laboratory setting when adult participants were trained in a rapid presentation paradigm on the 3 categories shown in Figure 5 (Ramscar et al., 2010). Each category comprised a series of "fribble" exemplars (James, Shima, Tarr, & Gauthier, 2005), and each fribble was made up of one of three solid body shapes, and a range of other features that can be seen as tails, legs, etc. Critically, the body shapes, which are highly salient features, were distributed systematically across the three categories, so that 75% of the members of one category and 25% of the members of another category shared the same body type. However, while body type was thus not discriminative category membership, the other features were. For example, as Figure 5 shows, although high-frequency *tobs* have the same bodies as low-frequency *deps*, the high-frequency *tobs* have tripod legs, whereas the low-frequency *deps* are bipedal. To successfully learn the categories, participants had to learn to ignore the uninformative body features and focus on, eg.g, the legs of the fribbles, a process that will be facilitated by cue competition, because the greater level of prediction errors generated by the body cues as compared to the other fribble features will lead to a shift in cue values from the bodies to the other features.

Learning the objects as cues to discrete labels, such as "wug" or "dax" (FL-learning) thus allows for competitive learning amongst the co-varying cues presented by the objects, enabling participants to learn to discriminate the informative features of the fribbles (e.g., legs) from those that were uninformative (the fribble bodies), and the performance participants given FL-training on a subsequent categorization test was very good: participants trained in this way were able to classify low and high frequency exemplars successfully. However, when the temporal arrangement of labels and fribbles was reversed, so that the process became one of learning to predict a set of features from a discrete label (LF-learning) things changed. In this situation, by virtue of there being only a single, non-varying cue to each trial – the label – it follows that competition *between* cues cannot occur; and in the absence of an information structure that facilitated the *unlearning* of uninformative dimensions in the category structures, participants trained with labels as cues to fribbles failed to learn to categorize the low frequency items.

WHEN MIGHT LABEL-TO-FEATURE-LEARNING BE USEFUL?

So far our comparisons of **FL**- and **LF**-learning, and their ability to promote discrimination learning (albeit in somewhat idealized conditions), have focused on learning meaningful relationships between words and the world. A language is, however, not just (or even) a series of binary mapping between forms and meanings. Linguistic messages are, for the most part, arranged sequentially, such that linguistic regularities (words, affixes, etc.) do not only serve to convey semantic information (reducing uncertainty about what is meant), but rather—to a greater or lesser degree—they also serve to convey grammatical information

(reducing uncertainty about the form of a message). This is most readily apparent in the case of function words (e.g., articles, which provide information about upcoming parts of speech), but even words that are apparently contentful can be seen to be playing a similar role. Take, for example, a sequence of English pre-nominal adjectives such as *cute little*... While traditional linguistic analysis tends to hold that pre-nominal adjectives modify the meaning of nouns, in any large corpus of English, *baby*, *puppy*, and *kitten* will be among the small set of nouns that are more likely to occur after *cute little*...

Given that only an extremely small percentage of babies, puppies, and kittens are not cute and little, adding *cute* and *little* to *baby*, *puppy*, and *kitten* appears to do little to nothing when it comes to modifying the meanings of these nouns. On the other hand, given that both common and proper nouns are highly diverse sets of entities, such that there are far more different noun types than for other parts of speech, it follows that, for both speaker and listener alike, the amount of information (measured as *entropy*, or uncertainty) that must be processed whenever a noun is encountered will be correspondingly greater than for other parts of speech. This information processing offers an alternative explanation for why people choose to mention cute and little prior to talking about babies, puppies, and kittens: given that information processing is an inherently deductive, discriminative process (Shannon, 1956), *cute* and *little* serve to increase the likelihood of the actual noun that a speaker or listener is about to encounter, because if speakers and listeners regularly mention cute and little prior to talking about babies, puppies, and kittens, then *cute* and *little* will serve to increase expectations regarding these nouns, and simultaneously decrease expectations for others.

Some support for this analysis comes from studies of grammatical gender. The agreement relations between articles and nouns in gender marking languages fall somewhere between those described for articles and pronominal adjectives in English (described above). Gender-marking languages makes use of multiple determiner classes which serve a role akin to that described for pronominal adjectives, and there is considerable evidence that native speakers use the discriminatory information provided by gendered articles to help reduce uncertainty about the specifics of upcoming linguistic material (Dahan, Swingley, Tanenhaus, & Magnuson, 2000; Grosjean, Dommergues, Cornu, Guillelmon, & Besson, 1994; van Heugten & Shi, 2009; Lew-Williams & Fernald, 2007; 2010; 2012). Consistent with this, in an artificial language learning study Arnon and Ramscar (2012) found that adults who successfully learned the relationship between two articles (*sem* and *bol*) and fourteen novel labels for familiar concrete objects (e.g., piano-*slindot*) subsequently showed better learning of the noun-labels and their meanings than participants whose training inhibited the learning of the relationship between the articles and noun-labels.

These considerations suggest that while *convergent* **FL**-learning might indeed be useful in learning to abstract semantic relations in language, *divergent*

LF-learning might have an important role to play in reducing grammatical uncertainties in linguistic processing. As Figure 6 shows, if we conceive of **FL**– and **LF**-learning in terms of the convergent and divergent relations between a label and a set of elements that they embody, then convergent (**FL**) relations will tend to facilitate cue competition between elements, and abstraction of the informative dimensions that best predict a label, whereas the divergent schema (**LF**) will facilitate learning of the probabilities of any elements given the label as a cue (Ramscar et al., 2010).

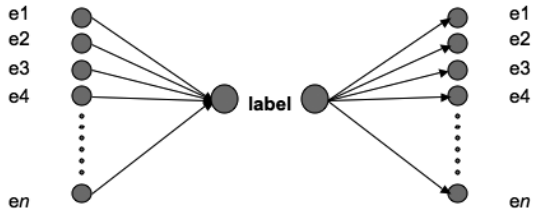

Figure 6. Sequential relationships between a linguistic regularity (label) such as a word or an affix and other elements in the world (or in a language). A convergent schema (**left**) will tend to facilitate cue competition between elements, and abstraction of the informative dimensions that best predict a label, whereas the divergent schema (**right**) will facilitate learning of the probabilities of the elements given the label (see also Ramscar et al, 2010).

To return to the function of prefixes and suffixes (and to treat the notion of prefix especially in its broadest sense, as a regularity that precededs a given linguistic word form), this in turn suggests a more nuanced explanation of their functional role has been previously suggested (see e.g., St Clair et al., 2009). Instead of suggesting that suffixes are more informative about grammatical categories per se, this analysis suggests that prefixes will tend to serve to reduce uncertainty about the form of upcoming material (i.e., they will be more informative about the members of a category), whereas suffixes will tend to serve to abstract a common dimension in that material, such that they will be more informative about the category itself. Critically, the different influence of convergent and divergent relations on cue competition suggest that learning may vary depending on the relations promoted by suffixes and prefixes.

If the number of prefixes that are associated with a given set of items were to increase, such that, say, two determiners were used alternately with a class of gendered nouns, the information provided by each prefix would remain the same (it would probabilistically cue each of the set of nouns at its occurrence frequency). In processing terms, all that would change is that the uncertainty associated with the occurrence of prefixes themselves would increase (to keep with the gender analogy, uncertainty about which gender marker one might expect at any given point would increase). However, if the number of suffixes

associated with a set of items were to increase (as happens with English verb inflections), cue competition will cause any coherent relations between the set of items and the occurrence of the suffixes to rise in prominence. Thus while increasing the number of prefixes will not alter the information a given prefix provides about a given set of items (albeit that the entropy of the prefixes will increase), the information suffixes provide about a set of items will change as the number of suffixes increases, because cue competition will cause each suffix to become more informative about a subset of the cues associated with the set of items and less informative about the overall set of cues associated with that set.

To empirically examine the different effects of convergent and divergent relations on affix learning, an auditorily presented novel language was created in order to contrast their effects on learning. The language comprised a set of novel noun labels (assigned to common English noun categories) and a set of affixes. Participants were first trained on the nouns, and then exposed to the language in an artificial "radio broadcast." After this, Participants were subjected to a series of tests to examine the predictions of this analysis: that suffixing would lead to better learning of abstract categorical relations between the nouns, whereas prefixing would lead to better learning of grammatical relations between the lexical forms and affixes.

ARTIFICIAL LANGUAGE LEARNING STUDY

Participants

35 native English-speaking undergraduates at Stanford University participated in the training study (20 F, 15 M). A further 138 undergraduates provided pre-ratings as part of a questionnaire packet.

Method

Participants were informed that in previous work, we had discovered that by teaching many people a few words of a language, we had been able to generate plausible translations by averaging across the responses of individual listeners who were unable to understand what they were listening to. For the purposes of the study, the current participants were told that they would learn a few words of a near extinct language, and then be asked to listen to a transcript of a radio broadcast in that language. When listening to the broadcast, participants were asked to not worry about trying to understand what they heard, but rather to focus on listening and attending to the speaker, since we were interested in what they would implicitly learn from doing so.

Vocabulary learning. In the first phase of the study, participants were required to learn a small set of sixteen vocabulary words. Each of the "foreign" words was assigned the meaning of a common English noun (see Table 1), and in training, one of a series of different pictures depicting each meaning appeared on a computer screen for 700ms, after which the participants heard its name pronounced while an orthographic representation of the name appeared to the right of the picture. 700ms after naming, the screen was blanked for 1000ms, and then the next training trial began. The words were presented in a random order for a total of 10 times each.

Table 1. The vocabulary of the artificial language employed in the training experiment.

Nouns and assigned meanings			
Gorok	bike		
Etkot	key		
Govom	boat		
Bahlot	fork		
Toonbot	clock		
Hertin	iron		
Jatree	pan		
Slindot	piano		
Hekloo	bath		
Pikroo	car		
Geesoo	hat		
Herdip	house		
Sodap	spoon		
Viltord	plane		
Panjol	television		
Fertsot	sock		
Prefixes			
itu;	oos;	mal;	poz
Suffixes			
sem;	bol;	lek;	dut
Nonsense strings			
Os ferpel een;	Slind vargt apt		

After training, pictures of each named object were again presented in a random order (without their names), and participants were asked to press a key marked with three letters corresponding to the beginning of each name on a keyboard. If the answer given was correct, the computer added the name to the screen in blue for 700ms. If the wrong answer was given, the computer beeped and added the correct name in red. If participants made errors, they were given a rest period, and then took another training session in which each name was presented three times in the same way as in the initial training session. Every participant was required to name every object correctly before moving onto the next part of the experiment.

"Language" learning. Once a participant had successfully passed the test in the learning phase, they were offered the chance of a break, and then asked to listen to the "radio broadcast". The broadcast consisted of a monolog utilizing the 16 trained words, 4 novel prefixes and 4 novel suffixes, and 2 nonsense strings. Across the broadcast, 8 of the trained words were presented with the same prefix and suffix every time they were spoken, 4 of the words were presented with the same suffix but alternated between two of the prefixes equally, and 4 of the words were presented with the same prefix but alternated between both suffixes equally (see Table 2). The prefix– novel-noun –suffix strings were then interspersed with the nonsense strings to create the "broadcast," of which there were 16 versions, across which the order of presentation of each of the different types of prefix– novel-noun –suffix types was counterbalanced. Each of the two nonsense strings comprised a three-word sequence, and a nonsense string was interpolated into the broadcast before two prefix– novel-noun –suffix strings, which meant that (treating the prefix and suffix markers as separate words), all of the novel-nouns were

heard in each 72 words, and all of possible permutations of the prefix– novel-noun –suffix strings were heard every 144 words (which took around a minute to present). Participants listened to a "radio broadcast" lasting approximately 15 minutes in which they heard every noun 30 times, and every possible prefix– novel-noun –suffix permutation at least 15 times. To create the broadcast, a British English speaker (MR) was recorded pronouncing each of the vocabulary items. The recording was tokenized, and the tokens were combined to form the strings presented in training.

Table 2. The full set of prefix– novel-noun –suffix permutations spoken across the broadcast. Words presented with two suffixes are shown in the top group, and words that appeared with two prefixes are shown in the middle group. Words with consistent pre– and suffixes are shown at the end.

oos	bahlot	lek
oos	bahlot	bol
oos	fertsot	bol
oos	fertsot	sem
itu	geesoo	sem
itu	geesoo	dut
poz	herdip	bol
poz	herdip	lek
mal	slindot	dut
mal	slindot	sem
itu	toonbot	lek
itu	toonbot	dut
mal	etkot	bol
poz	etkot	bol
mal	govom	dut
poz	govom	dut
itu	hartin	dut
oos	hartin	dut
itu	jatree	bol
oos	jatree	bol
mal	pikroo	sem
oos	pikroo	sem
poz	viltord	lek
itu	viltord	lek
oos	gorok	bol
mal	hekloo	dut
itu	panjol	lek
poz	sodap	sem

Importantly, in the course of hearing the materials, all of the participants were exposed to items where:

1. The prefix was always consistent but where the suffix varied;
2. The prefix varied and the suffix was consistent;
3. Both the prefix and suffix were consistent.

This allowed participants' ability to accurately discriminate the actual set of affixes heard from those that were not heard to be tested, as well as the effects of affix-type variance on the discrimination of categorical information to be explored.

Testing. Learning was tested in two ways:

1. Participants were presented with pictures of the objects associated with the content words and asked to rate how similar they thought the objects were on a scale of 1 to 10 (1=not similar). Object pairs were presented in groups of 4 (i.e., 8 objects appeared on each page of the test booklet), and participants were asked to make their scorings reflect any relative similarity differences that they perceived across the pairings. (138 control participants also competed this task: each of these participants rated 4 pairs of objects without undergoing any training.)
2. Participants completed a forced choice task in which they either heard strings that had been presented in training, along with one of the objects associated with the string, or "lure" strings which had not been encountered in training. Participants were asked to judge whether the strings were *old* (had been encountered in training) or new (had not been encountered in training). Three types of lures were used: (1) those in which the suffix and object-label were consistent with training, but one of the prefixes was inconsistent; (2) those in which the prefix and object-label were consistent with training, but one of the suffixes was inconsistent; and (3) those in which the suffix and prefix were consistent with training, but one of the labels was inconsistent with a depicted object (Figure 7).

1. *Os ferpel een mal pikroo sem* **versus** *Os ferpel een itu pikroo sem*
(correct prefix) (incorrect prefix)

2. *Slind vargt apt oos bahlot lek* **versus** *Slind vargt oos bahlot sem*
(correct suffix) (incorrect suffix)

3. *Os ferpel een poz viltord lek* **versus** *Os ferpel een poz herdip lek*
(correct semantics) (incorrect semantics)

Figure 7. Examples of correct items and lures used in testing. (1) Suffix and object-label are consistent with training, but one of the prefixes is inconsistent; (2) the prefix and object-label are consistent with training, but one of the suffixes is inconsistent; (3) the suffix and prefix are consistent with training, but one of the labels is inconsistent with the picture.

Results

The average performance for each task when either the suffix was consistent and the prefix alternated between two forms or else the prefix was consistent and the suffix alternated between two forms plotted in Figure 8. As can be seen, whereas prefix consistency resulted in greater grammatical accuracy than suffix consistency ($t(31)=6.39, p<0.0001$), objects whose labels were paired with a consistent suffix were later judged to be (relatively) more similar than objects paired with a consistent prefix ($t(31)=2.2$, $p<0.001$; a repeated-measures ANOVA confirmed the interaction between affix consistency and test type).

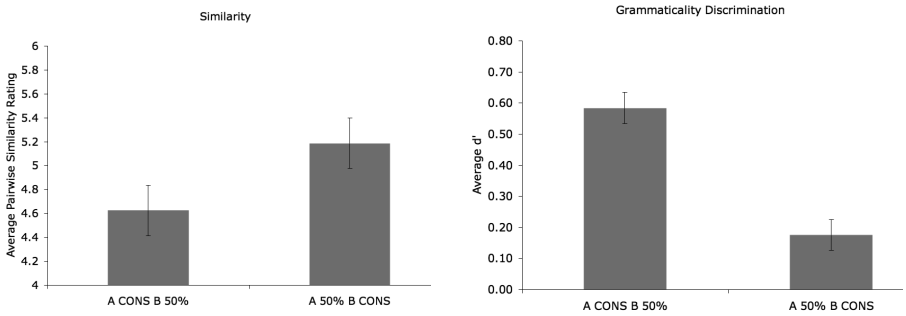

Figure 8. Mean performance in each test condition (similarity ratings left panel; grammaticality performance (d') right panel) when the prefix (A) was consistent and the suffix (B) alternated between two forms (left bar in each panel), and when the prefix (A) alternated between two forms and the suffix (B) was consistent (right bar in each panel).

Planned comparisons revealed that all the training conditions increased the pairwise similarity of the test items over the baseline (M=4.11): prefix consistent (M=4.65, t(306)=1.68, p=0.09, suffix consistent (M=5.18, t(306)=3.496, p<0.0001), prefix and suffix consistent (M=5.39, t(306)=5.675, p<0.0001). Similarity ratings for the prefix and suffix consistent control items was not significantly greater than for the items where only suffixes were consistent (t(94)=0.75, p>0.45), however when only prefixes were consistent, items were rated less similar on average than for the controls (t(94)=2.89, p<0.005).

In the grammaticality task, prefix consistent training led to better discrimination of lures where the suffix and object-label were consistent with training, but the prefix was inconsistent (t(31)=5, p<.0001), and also where the suffix and prefix were consistent with training, but the object was paired with an inconsistent label (t(31)=2.88, p<.01). However, participants' ability to discriminate items where the prefix and object-label were consistent with training, but the lure suffixes were inconsistent did not differ significantly when the suffixes were the only consistent affixes in training as compared to when the prefixes were the only consistent affixes (t(31)=1.22, p>.2; see Table 3 for mean Hit and False Alarm rates by item type).

Table 3. Hit and False Alarm rates on the grammaticality task by training and test type (figures represent the probability of each response type).

	Prefix		Suffix		Noun	
	Hit	False Alarm	Hit	False Alarm	Hit	False Alarm
A CONS B 50%	0.97	0.16	0.72	0.66	0.77	0.39
A 50% B CONS	0.88	0.69	0.84	0.63	0.88	0.75

DISCUSSION

Previous research has shown that convergent relations between linguistic labels and other elements tend to promote cue competition, whereas where these relations diverge, the resulting information structure inhibits cue competition (Ramscar et al., 2010). This (along with some very basic mechanisms of error-driven learning) predicts that suffixes and prefixes ought to promote different kinds of learning, with prefixing tending to promote the learning of the probabilities of any following elements in a sequence that follow a label, whereas suffixing will tend to promote the abstraction of common dimensions from a set of preceding elements. The results of the artificial language learning experiment broadly confirm these predictions. When words associated with common nouns were learned with consistent prefixes, participants learned the relationship between the prefixes and the noun labels, and the relationship between the noun labels and the objects associated with them, better than when words were learned with consistent suffixes. On the other hand, learning words associated with common nouns with consistent suffixes resulted in participants treating similarly suffixed nouns as being more similar than nouns learned with consistent prefixes (albeit that both kinds of training increased similarity above baseline). Or, to put it another way, it appears that while prefixes tended to make items more predictable (and made veridical discriminations easier), suffixes tended to make items cohere more, increasing the similarities between them.

It is important to acknowledge at this point that the treatment of prefixes in this paper has been somewhat at odds with many other analyses of affixing (see e.g. Sapir, 1921). Indeed, I have not explicitly dealt with the kind of prefixing that actually distinguishes prefixes and suffixes in traditional analyses. This is because my goal has been to outline and explore some very general constraints that information structure and learning mechanisms appear to impose on the form and content of language learning, rather than to account for the facts of a particular language. In earlier work (Ramscar et al., 2010; Arnon & Ramscar, 2012), my colleagues and I have presented evidence that where language learning occurs in semantic contexts, elements that may be described in hindsight by linguists as prefixes and suffixes are initially likely to be treated equivalently. Initially, this will also allow for meanings to be abstracted from prefixes in much the way that I have described for suffix learning above, such that the subsequent unlearning of semantic cues to prefixes will depend ultimately on the distribution of semantic cues, prefixes, and other lexical items (Figure 9): i.e., to some degree, *all* words must compete for meaning over time, regardless of their sequential relations, and the outcome of this competition will be determined by the degree to which semantic features (i.e., the world) are informative about a given word (Ramscar, Dye, & Klein, 2013); the results presented here, along with the earlier results reviewed, suggest that sequencing will influence the way in which informative relations are learned, and the ultimate shape of the distribution of forms and semantics in any given language.

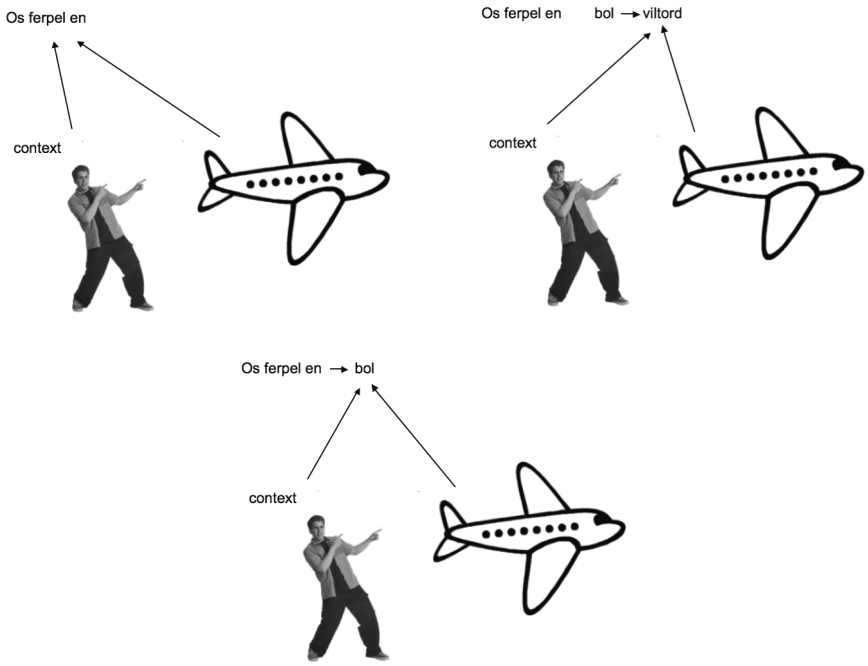

Figure 9. The cue structure of training trials in a study by Arnon & Ramscar (2013). Visual stimuli (the man and the plane) were present throughout each trial, while aural stimuli unfolded in time (the text representation of this is included as an illustration). Initially, each noun depiction (in this instance the plane) and each trial context (the gesturing man) are available as cues to the initial phrase "os ferpal en" (left). The initial phrase then serves as an additional cue to "bol" (center), which in can serve as an additional cue to "viltord" (right). As learning proceeds further, these cues will incrementally strengthen and weaken depending on the way that "os ferpal en," "bol" and "slindot" co-occur with each other, and with objects and events in the world. If these patterns of co-occurences vary, then any cues available in the linguistic and semantic environment will begin to compete for relevance in predicting "os ferpal en," and then "bol" and then "slindot." The value of the semantic and linguistic cues that a learner acquires for these elements (their meanings) will be determined by this competition.

Or, to put it another way, the reality of language learning is, of course, not as clear-cut as in the artificial task examined here (see Arnon & Ramscar, 2012), and patterns of divergent and convergent relations themselves will depend not only on relations between words, but on systematic relationships between the words of a language and their relationship to the world and to human experience (such that whether a given linguistic element should be seen as a prefix, a suffix or a compound is ultimately likely to be dependent on the distribution of a language, and the experience of a community of speakers). It is within systems of language that divergent and convergent relations (and suffixes and prefixes) make their different contributions to uncertainty reduction in language processing, and these contributions likely reflect a set of more general principles, which

constrain the way that languages exploit sequential order and learning to manage uncertainty in communication (Ramscar & Baayen, 2013). It seems likely that further study of the relationship between language structure, sequencing, and the interactions between these factors and the mechanisms through which humans learn, can help shed light on the way that language works, and help explain why languages have the structures that they do.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant Nos. 0547775 and 0624345 to MR. Petar Milin and Harald Baayen provided insightful comments on an earlier draft of this paper, for which I offer heartfelt thanks.

REFERENCES

Arnon, I., & Ramscar, M. (2012). Granularity and the acquisition of grammatical gender: How order-of-acquisition affects what gets learned. *Cognition*, *122*(3), 292–305.

Cutler, A., Hawkins, J. A., & Gilligan, G. (1985). The suffixing preference: A processing explanation. *Linguistics*, *23*, 723–758.

Dahan, D., Swingley, D., Tanenhaus, M. K., & Magnuson, J. S. (2000). Linguistic gender and spoken-word recognition in French. *Journal of Memory and Language*, *42*(4), 465–480.

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

Daw, N., & Shohamy, D. (2008). The cognitive neuroscience of motivation and learning. *Social Cognition*, *26*(5), 593–620.

Dickinson, A. (1980). *Contemporary animal learning theory*. Cambridge: Cambridge University Press.

Greenberg, J. H. (1957). Order of affixing: A study in general linguistics. In J. H. Greenberg (Ed.), *Essays in linguistics* (pp. 89–94). New York: Wenner-Gren Foundation for Anthropological Research.

Grosjean, F., Dommergues, J. Y., Cornu, E., Guillelmon, D., & Besson, C. (1994). The gender-marking effect in spoken word recognition. *Perception & Psychophysics*, *56*(5), 590–598.

Hawkins, J. A., & Cutler, A. (1988). Psycholinguistic factors in morphological asymmtery. In J. A. Hawkins (Ed.), *Explaining language universals* (pp. 280–317). Oxford, England: Basil Blackwell.

Hawkins, J. A., & Gilligan, G. (1988). Prefixing and suffixing univerals in relation to basic word order. *Lingua*, *74*, 219–259.

Hupp, J. M., Sloutsky, V. M., & Culicover, P. W. (2009). Evidence for a domain-general mechanism underlying the suffixation preference in language. *Language and Cognitive Processes*, *24*, 876–909.

James, W. (1890). 1950. *The principles of psychology*, New York: Dover Publications.

Kamin, L. J. (1969). Predictability, surprise, attention, and conditioning. In B. Campbell & R. Church (Eds.), *Punishment and aversive behaviour* (pp. 279–298). New York: Appleton-Century-Crofts.

Kruschke, J. K. (2011). Models of attentional learning. In E. M. Pothos & A. J. Wills (Eds.), *Formal Approaches in Categorization* (pp. 120–152). Cambridge: Cambridge University Press.

Lew-Williams, C., & Fernald, A. (2007). Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science*, *18*(3), 193–198.

Lew-Williams, C., & Fernald, A. (2010). Real-time processing of gender-marked articles by native and non-native Spanish speakers. *Journal of Memory and Language*, *63*(4), 447–464.

Grüter, T., Lew-Williams, C., & Fernald, A. (2012). Grammatical gender in L2: A production or a real-time processing problem. *Second Language Research*, *28*(2), 191–215.

McLaren, I. P. L., & Mackintosh, N. J. (2000). An elemental model of associative learning: I. Latent inhibition and perceptual learning. *Animal Learning & Behavior*, *28*, 211–246.

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

Niv, Y. (2009). Reinforcement learning in the brain. *Journal of Mathematical Psychology*, *53*(3), 139–154.

Osgood, C. E. (1949). The similarity paradox in human learning: A resolution. *Psychological Review*, *56*, 132–143.

Pavlov, I. P., & Anrep, G. V. (1927). *Conditioned refle*xes. New York: Courier Dover Publications.

Ramscar, M., & Baayen, H. (2013). Production, comprehension and synthesis: A communicative perspective on language. *Frontiers in Language Sciences*. *4*:233. doi: 10.3389/fpsyg.2013.00233

Ramscar, M., Yarlett, D., Dye, M., Denny, K., & Thorpe, K. (2010). The Effects of Feature-Label-Order and their implications for symbolic learning. *Cognitive Science*, *34*(6), 909–957.

Ramscar, M., Dye, M., Gustafson, J. W., & Klein, J. (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, M., Dye, M., Popick, H. M., O'Donnell-McCarthy, F. (2011). The enigma of number: Why children find the meanings of even small number words hard to learn and how we can help them do better. *PLoS ONE, 6*(7): e22501. doi:10.1371/journal.pone.0022501

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

Rescorla, R. A. (1988). Pavlovian conditioning: It's not what you think it is. *American Psychologist*, *43*(3), 151–160.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.

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.

Santos, L. R., Flombaum, J. I., & Phillips, W. (2007). 15 The Evolution of Human Mindreading: How Nonhuman Primates Can Inform Social Cognitive Neuroscience. *Evolutionary cognitive neuroscience,* 433–456.

Sapir, E. (1921). *Language*. New York: Harcourt, Brace and Company.

Schultz, W., & Dickinson, A. (2000). Neural coding of prediction errors. *Annual Review of Neuroscience*, *23*, 473–500.

Schultz, W. (2006). Behavioral theories and the neurophysiology of reward. *Annual Review of Psychology, 57*, 87–115.

Schultz, W., Dayan, P., & Montague, R. R. (1997). A neural substrate of prediction and reward. *Science*, *275*, 1593–1599.

Shannon, C. E. (1956). The Bandwagon. *IRE Transactions on Information Theory*, *2*(1), 3.

Van Heugten, M., & Shi, R. (2009). French learning toddlers use gender information on determiners during word recognition. *Developmental Science*, *12*(3), 419–425.

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

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