Michael Ramscar and Robert Port (2015). Categorization (without categories). In, E. Dawbroska & D. Divjak (Eds.), Handbook of Cognitive Linguistics. De Gruyter Mouton. 75-99

4. Categorization (without categories)

1. Introduction

Cognitive linguists view language as a social artifact shaped by learning and cultural transmission, and emphasize the role of *categorizatio*n in shaping our linguistic capacities (Lakoff 1987; Taylor 2003). This has resulted in something of a division of labor, as linguists seek to explain the role of categorization in shaping the functional properties of language, while psychologists seek to uncover the cognitive bases of categorization itself.

These endeavors share many assumptions and questions: how do people decide that an aspect of a scene should be labeled (in English) *a mountain* or *tree*? How do they determine that an instance of speech contains the sound [ɒ] that distinguishes *water* from *waiter*? And both approaches assume that the fact that people use words like *water* and *waiter* indicates they have access to the *concepts waiter*, *water*, and *[ɒ]*. These concepts are discrete mental units that (somehow) specify their content, and can be combined with other concepts to create thoughts, utterances and sentences.

Although this assumption makes some intuitive sense, it not clear that English speakers' use of the word *tree* does warrant the assumption that each speaker possesses a coherent, unified representation of the concept *tree*, or that trees form a coherent class of natural objects (Wittgenstein 1953; Quine 1960). Moreover, the struggles that are evident when categorization researchers seek to define the object of their study suggest that these assumptions may be unwarranted:

> The concept of concepts is difficult to define, but no one doubts that concepts are fundamental to mental life and human communication. Cognitive scientists generally agree that a concept is a mental representation that picks out a set of entities, or a category. That is, concepts refer, and what they refer to are categories. (Rips et al. 2013: 177)

However, because *reference* and *representation* are as ill defined as *concept*, describing concepts as mental representations that refer to classes of entities in the world simply exchanges one poorly defined term for another. This problem led Wittgenstein (1953) and Quine (1960) to reject the idea of concepts as discrete constructs, and emphasize instead the way that words function in systems, and the way that meanings result from the way words are *used* in these systems.

This distinction − between concepts as discrete mental tokens, and concepts as emergent aspects of systems − is usually glossed over in the literature, but its implications for language, both in terms of both semantics and linguistic categories themselves, are far-reaching. To establish which characterization better describes the results of many years of empirical study, this chapter reviews the results of this work, along with the many computational models that have been developed to account for them, and recent work seeking to match these models to neural structures in the brain.

The findings of these lines of work make it clear that minds do not learn inventories of discrete, stand-alone concepts. Instead, human conceptual capabilities are systematic: they are the product of a rich capacity to discriminate and learn *systems* of alternative responses (behaviors, affordances, words, etc.) and to use them in context. From this perspective English speakers do not acquire discrete concepts of *tree* or *friend*, but rather they learn a system of linguistic contrasts, and how to discriminate when to use the words *tree* (rather than *bush*, *oak* or *shrub*) or *friend* (rather than *buddy* or *pal*) in order to satisfy their communicative goals. We conclude this review by briefly describing what this implies for future directions of research in cognitive linguistics.

2. Concepts, categories and categorization

2.1. Concepts and labels

It seems clear that the existence of individual nouns need not entail the existence of corresponding individuated cognitive entities, yet speakers tend to talk about *concepts* as if they are discrete, countable things. Thus dictionaries characterize *concepts* as "directly intuited objects of thought" or "ideas of things formed by mentally combining their characteristics" (passing the buck of defining *concept* onto *idea* and *object of thought*). This practice extends to the categorization literature, which focuses on either discrete, artificially defined concepts, or the nature of frequent linguistic items like *tree*.

From this circumscribed perspective, researchers seek to explain how cognitive representations of concepts can be described in terms of relations between *features* (and where a *category* is the set of instances that exemplify a *concept*). To work, this approach to requires clear definitions of what concepts and features are, and this, as noted above, is problematical: if the only thing that unifies the category of *games* is the fact that we call them games − i.e., they share the feature of being called *game* in English − then a definition of the concept *game* will add no more to our understanding of concepts than saying that *games* are whatever English speakers call *games*. Further, as Wittgenstein (1953) observed, if we take the common English word *game*, then *baseball* is a game, whether played by children for fun or by professionals for their livelihood; *polo* and *hopscotch* are games, as are *scrabble*, *solitaire*, and *monopoly*; yet *stockbroking* is not usually called a game, nor are *proofreading* nor *cavalry-charges* (military *maneuvers* are called games in peacetime but not wartime). And although many local similarities can be seen between different games, there are equally many local similarities between games and non-games (Goodman 1972), such that it appears that the only things that joins *games* together while ruling out *non-games* is the fact that some things are conventionally called *games* while others are not (Ramscar and Hahn 1998).

The circular relationship between labels and concepts is not restricted to abstract words like *game*: the English word *tree* does not correspond to a coherent class of objects, even leaving aside uses like *tree diagram*. Biologically, pine trees are *gymnosperms* (conifers), whereas oak trees are *angiosperms* (flowering plants). Oaks share an evolutionary lineage with cacti, daisies and roses; pines belong to a more primitive genus (Lusk 2011). *Vegetable* is a similar pseudo-natural kind: English *tomatoes* are vegetables, but most other edible fruits are not (Malt et al, 2010). And while even natural concepts (in the colloquial sense of the word concept) seem arbitrary because they lack defining features, defining features raise problems in turn that are disturbingly similar to those posed by concepts: What are they? Do features themselves merely reflect labels, etc. (Ramscar et al. 2010b; Port 2010; Port and Leary 2005)?

To illustrate the subtle problems these issues pose when it comes to forming a clear understanding of the role of concepts in categorization, consider a rare attempt to clearly define the terms concept and category from the literature: Goldstone and Kersten (2009) describe concepts as mental representation of individuals or classes that specify *what* is being represented and how to categorize it. They then note that if a concept is a mental representation of a class, and a category a set of entities that are appropriately categorized together by that concept, this leaves the question of whether concepts determine categories or categories determine concepts open:

> If one assumes the primacy of external categories of entities, then one will tend to view concept learning as the enterprise of inductively creating mental structures that predict these categories. […] If one assumes the primacy of internal mental concepts, then one tends to view external categories as the end product of applying these internal concepts to observed entities. An extreme version of this approach is to argue that the external world does not inherently consist of rocks, dogs, and tables; these are mental concepts that organize an otherwise unstructured external world. (Goldstone and Kersten 2003: 600).

However, the basis for these distinctions is dubious: It is demonstrably the case that the world does *not* inherently consist of rocks, dogs, and tables. *Dog* and *table* are English words (i.e., culturally defined labels), and there is little evidence to support the idea that natural partitions in the universe are constrained by the vocabulary of English. Indeed, as *tree* and *vegetable* illustrate, there is little reason to believe the English lexicon maps onto reality in a privileged way at all; *trees* are simply what English speakers call trees.

So what is a concept, and how do categories relate to concepts? The answer is that in practical terms, regardless of what researchers take the relationship between words and reality to be, concepts are determined by the ways things are labeled. A rock is an instance of *rock* if English speakers would call it a rock, and when researchers talk about two items as instances of a concept they simply mean that the items can be spoken about using the same linguistic symbol (and, as Gahl 2008 shows, even apparent homophones are symbolically distinct).

This point even extends to the artificially defined concepts used in experiments, because the features used in their definitions are themselves linguistic concepts like red, square, etc. Indeed, even where concepts are only implicitly labeled in a study's procedure − for example in pigeon experiments, or in abstract discrimination tasks − the relevant stimulus dimensions will have been explicitly labeled at some point in its design (e.g., Zentall et al. 2014; Billman and Knutson 1996).

In other words, irrespective of whether researchers believe in the primacy of the word or the primacy of the world, in practice they study concepts that are determined by labels. For clarity, in the rest of this review the word *concept* will be used to describe a specific relationship between a group of items: that they share a label, typically a word (or phrase, discriminable symbol, etc.).[1] We will use *concept learning* to describe the way the relationship between a set of items and a label is learned: That is, the process by which people learn whatever it is that enables them to respond to new items in a manner appropriate to a label. *Category* will be used to describe a set of items with a common label (including new items that could be considered to be members of that category in an appropriate context), and *categorization* will be used to describe the various things people are asked to do with category members, i.e., sort them, label them, make inferences about them, etc.

2.2. One label; Two ideas

Many of the confusions that abound in discussions of categorization research arise out of the fact that it comprises two distinct lines of study:

1. *Concept learning* experiments originating in the associative paradigms that dominated early psychology, and which focus on classification and response discrimination tasks, usually employing artificial concepts;
2. Studies of *the structure of natural language concepts* that measure people's behavior as they respond to uses of words and phrases.

---

[1] This does not means that people only learn in reference to labels, or that they only acquire "conceptual knowledge" that can be labeled. Rather, for obvious reasons, there is little discussion of completely non-verbalizable content in the categorization literature.

Because these two approaches employ the same terminology, but embody very different methodological approaches, potential for confusion abounds in the literature, especially when, as is common, results from studies of artificial concepts are discussed in the same breath as natural concepts.

2.3. Concept learning: from rules and definitions to prototypes and exemplars

Studies of *concept learning* typically examine people's ability to discriminate the appropriate dimensions of stimuli and learn to match them to discrete responses (Hull 1920; Smoke 1932). Researchers examine questions such as whether associations increase gradually or are better characterized in all-or-none terms (Trabasso and Bower 1968) and whether learning conjunctive dimensions (e.g., *blue* AND *triangular*) is easier than disjunctive dimensions (e.g., *blue* OR *triangular*; Bruner et al. 1956; Shepard et al. 1961).

Because artificial concepts are defined by feature-combinations, researchers often equate concept learning with acquiring a rule defining some kind of membership criterion (e.g., *"rule-based" concepts*; Smoke 1932): concepts are descriptions of the appropriate dimension(s) for class inclusion and categorization is a process in which item features are matched to rules across an inventory of concepts. Rule-based concept learning thus resembles early speculations about word meanings, e.g., Frege's (1892) distinction between the *intensions* and *extensions* of concepts (which is still widely used in linguistic analyses today): A concept's intension is the set of attributes defining its members, while its extension comprises its actual members.

Thus the intension of *bachelor* might include characteristics such as male, unmarried and adult, making its extension the set of male, unmarried adults in the world. And this would mean that both the Pope and an unmarried man cohabiting with the same partner for 25 years are bachelors. One can, of course, amend the intension of bachelor to exclude Popes and cohabitees to fix this, but what is important to note here is a critical *difference* between natural language and artificial concepts: the latter are whatever a researcher defines them to be, whereas definitions for natural language concepts can only be imputed.

It follows that, theoretically, the question of whether there is an appropriate conceptual definition for bachelor at all is equally as valid as what the conceptual definition of bachelor is. Or, to put it another way, although the definitional status of artificial concepts can be taken for granted, valid definitions for natural concepts may not actually exist (Wittgenstein 1953). Intriguingly, this latter notion is supported by the results of artificial concept learning studies themselves: Findings from numerous experiments indicate that people don't actually learn to represent rules or feature specifications for carefully defined concepts even when they encounter them (Sakamoto and Love 2004).[2]

For example, Posner and Keele (1970) showed that people are better at classifying previously unseen *typical* artificial category exemplars than less typical exemplars they

---

2 Since "features" in artificial concepts are natural concepts at another level of abstraction, these findings are not entirely surprising.

have actually seen in training. Along with numerous other similar findings, this result suggests that people learn *prototypical* information about item-categories, such that even well defined concepts are not learned as definitions.

Other findings muddy the waters still further. When typicality differences are controlled for, participants *are* better at categorizing items seen in training than new items (Nosofsky 1992; Smith and Minda 2000). Similarly, less typical new items similar to items seen in training are categorized more accurately and quickly than typical new items that are not (Medin and Schaffer 1978). These results suggest that participants learn details about the specific items they are exposed to (i.e., exemplars) rather than abstracting rule-based representation or pure prototypes.

On the other hand, concept learning does result in some abstraction: Posner and Keele (1967) showed that although participants retain information about the letters *a* and *A* for a couple seconds, these initial encodings give way to representations in which *a* and *A* are stored as exemplars of a more abstract (case-invariant) letter name. Similarly although participants adjusting the length of a reference line to that of a Müller-Lyer stimulus which is either in view or in memory exhibit a pattern of bias consistent with the Müller-Lyer effect in both cases, the adjustments made from memory are further biased towards the average line length presented in the experiment (Crawford et al. 2000).

What people learn about concepts is further influenced by the tasks they perform in experiments (Love 2002). Learning in inference tasks (answering questions like, "This is a mammal. Does it have fur?") highlights dimensions that are typical rather than diagnostic of concepts (which in this case would involve milk-producing glands). By contrast, classification tasks ("This has milk-producing glands. Is it a mammal?") promote learning of the diagnostic information that discriminates between categories (Yamauchi et al. 2002).

Finally, Brooks et al. (1991) have shown how specific exemplars play a role in experts' use of well-defined concepts. For example, doctors often base diagnoses on recent cases, rather than more general abstractions, suggesting that expertise involves the acquisition of relevant exemplar knowledge, as well as rules.

This body of results is incompatible with the idea that concepts are defined in memory by stable feature sets, or that such things are plausible as theoretical postulates (Ramscar et al. 2013d, 2014). Even where people learn clearly specified concepts, they learn both more and less than a definitional account might imply, and what they learn is influenced by the tasks they perform during learning, by context, and by the process of learning itself (Arnon and Ramscar 2012; Ramscar et al. 2013d).

2.4. The structure of natural language concepts and the basic level

The other line of categorization research examines the knowledge associated with the words used in languages (Rosch and Mervis 1975; Rosch et al. 1976). Rosch and colleagues argued that everyday concepts are not structured in ways that reduce to definitions based on necessary and sufficient features. Although a given feature might be common to many items corresponding to a word's usage (e.g., *birds fly*), it might not be common to all (*penguins*) and it might also be common to items labeled using other words (*insects*). Rosch et al. proposed that natural concepts have a *family resemblance* structure, and that category membership (labeling) depends on similarities between members of a category.

An idea that naturally follows from this suggestion is that there are better and worse examples of a concept: Category members that share more properties with other members should better exemplify a concept than those sharing fewer properties. And studies confirm that people believe canaries are better examples of *birds* than *penguins* (Rosch and Mervis 1975), and that these goodness judgments correlate with the number of features that a given example shares with other examples.

Rosch et al. (1976) argue that the distribution of features among concepts results in natural clusters that maximize within-category similarity and minimize between-category similarity. They termed these basic-level concepts. Examples would be *dog* (as opposed to *dachshund,* or *pet*) and *house* (as opposed to *duplex*, or *mansion*). Rosch et al suggest that basic-level categories are (a) preferred by adults in naming objects in tasks that contrast various levels of abstraction (b) used more in child directed speech, (c) learned first by children, and (d) are associated with the fastest categorization reaction times.

Although Rosch et al. repeatedly show that people are more likely to use basic level words than those at other levels in their abstraction hierarchy, they paradoxically maintain this is *not* because basic level words are more frequent (presenting evidence from small written corpora in support of this idea). However, it is worth noting first that Rosch et al. acknowledge that "basic level" categories can be influenced by culture and expertise (thus, for a real-estate agent, *colonial* may be a basic-level concept), and second, that word frequency effects are ultimately conditioned on an individual's experience, not corpus statistics (Ramscar et al. 2014). Further, the basic level labels studied by Rosch et al. are high frequency English nouns. Because of this, it is unclear whether basic level categories should be seen as offering insight into the structure of the world, personal and cultural structures, or interactions between the two (Malt et al. 2010).

Work in this tradition poses another problem for discrete theories of concepts because it provides evidence that some − if not all − natural language categories lack clear boundaries. Labov (1973) showed that there is a great deal of variability in the way people use terms such as *cup*, *bowl*, etc., with different labels being assigned to the same containers both between speakers and within speakers depending upon the linguistic context. If people are asked to look at a picture of an object whose shape is half way between a (tea) *cup* and a (soup) *bowl* and told that it contains mashed potatoes, they tend to consider the object to be a *bowl*. But if the ambiguous object contained hot coffee, it would tend to be considered a *cup*. Similarly, in a study of exemplar-category pairs (e.g., *apple-fruit* or *chair-furniture*) McCloskey and Glucksberg (1978) found not only substantial between- and within-participant disagreement on category membership (measured over successive test-sessions) but also showed that levels of disagreement correlate with independently derived typicality ratings: McCloskey and Glucksberg's participants were certain that *chair* belonged to the category *furniture*, and that *cucumber* did not. However, there was much disagreement as to whether a *bookend* belonged to the category *furniture*, with many participants differing in their judgments from one session to the next.

Categorization in one domain can also be influenced by information from another. For example, thinking about space can influence subsequent temporal categorization judgments: the question *Next Wednesday's meeting has been moved forward two days:*

what day is the meeting on now? is more likely to be answered with Friday than Monday by people who have been encouraged to think about moving towards a physical destination rather than staying in one place (Boroditsky and Ramscar 2002; Evans this volume). Rating the sensibility of fictive motion sentences (e.g., *Seven trees ran along the driveway* vs. *There were seven trees along the driveway*) also produces a similar, predictable influence on temporal categorization (Ramscar et al. 2010a). However, although time and space seem to be systematically linked, the basis of this linkage ultimately appears be lexical: the patterns of priming observed in these experiments are highly consistent with patterns of lexical co-occurrence in English (Ramscar et al. 2010a).

Finally, we should note that Rosch (1978) argued that it would be a mistake to assume that the discovery of prototype effects indicated that word meanings are themselves represented by prototypes. Yet the idea that concepts *can* be assumed to have prototypical representations has since been proposed in various guises: as *frames* (e.g., Fillmore 1982); as *Idealized Cognitive Models* (ICMs, Lakoff 1987); as *image schemas* (Johnson 1987); and *domains* (Lakoff 1993). It is thus worth stressing that none of these suggestions make it clear what is or is not part of a specific frame, ICM, or domain, or indeed how concepts are actually represented in terms of these constructs. Thus despite the fact that these suggestions are often referred to as theories of representation, they are more akin to the phenomenological descriptions that Rosch suggested that prototypes are than theories of conceptual representation (Cienki 2007; these points also apply to *ad hoc categories*, Barsalou 1983).

2.5. Two Traditions – One Conclusion

Results from both lines of categorization research support the conclusion that words are *not* associated with invariant context-independent definitions. Even the learning of well-defined concepts appears to be sensitive to a range of contextual factors, such that people learn context-sensitive representations of even rule-based artificial concepts. It appears that natural language concepts do not have stable structures within or between individual speakers in a community (Ramscar et al. 2013d), and that people's use of words (and phrases) reflects conventions that probabilistically govern language use in context. That is, while it is clear that people learn and use conceptual knowledge in systematic ways, the results reviewed here offer little support for the idea that this behavior relies on or reflects their possession of discrete representations of concepts.

3. Computational models of categorization

Numerous computational models have been proposed to account for the empirical results discussed above, allowing theoretical proposals concerning conceptual representations to be evaluated by formal simulations of behavior. The development of categorization models has been a particular feature of the artificial concept learning literature, in part because the controlled nature of artificial stimuli is more amenable to formalization than the study of everyday concepts based on social convention. However, one of the earliest and most influential computational models of categorization (Collins and Quillian 1969) is an outlier in that it sought to formally characterize everyday semantic knowledge.

3.1. Hierarchical semantic networks

The Collins and Quillian model proposes that word use reflects a hierarchical network in which stimulus properties are stored in memory according to their generality or specificity in relation to a set of related concepts. This postulates, for example, a taxonomic representation of animal knowledge where properties general to all animals such as *breathing* are stored at the top of the hierarchy with the concept *animal*. Properties generally true of fish are stored at the *fish* node, and general bird properties are stored at the *bird* node. Properties distinctive to individual sub-kinds (e.g., *robin*) are stored with the specific concept nodes they characterize (e.g., the property *red-breasted*). In this model, category membership can then be defined in terms of the positions of nodes in the hierarchical network. Many properties of each category can be read off from its position. Thus the node for *salmon* does not directly store the information that salmon are *animals*, since that fact is specified by the hierarchical connection between the *salmon*, *fish* and *animal* nodes.

However the Collins and Quillian model is not a straightforward inheritance model as these are commonly understood in computer science: this is because sub-kinds on lower nodes do not always inherit all the properties of higher nodes. For example, *can fly* is associated with the *bird* node − because flying is usually a distinctive property of birds − and exceptions to this feature (i.e., *penguin*) are stored at a sub-kind node for *does not fly*. Thus while it is often reported that increases in network distance between concepts and properties successfully predict the time it takes people take to verify that concepts have a given property (e.g., people verify that a *canary* is *yellow* faster than that it *has feathers*), given the lack of any principles specifying exactly *where* in the hierarchy a given feature is represented positively or negatively, it is more accurate to say that Collins and Quillian's intuitions about the relationship between the various words used for nodes and features correlated well with the behavior of their participants. (Nb. to some extent this criticism applies to *all* models that describe formal relationships between sets of arbitrary conceptual features defined by modelers.)

3.2. Prototype models

Prototype models are characterized as seeking to formalize concept representations in which degree of fit to the category is evaluated for the purposes of categorization. A *prototype* represents information about all relevant dimensions of the items in a stimulus set but the information is presented as some kind of average value across all exemplars. In a prototype model, a novel item is classified as a member of the category whose prototype it is most similar to (e.g., Hampton 1995). The values used to define the prototype for each category are updated when new examples are encountered. These models thus seek to capture the critical structure of a category, without having to encode every detail of every item that a participant has seen.

Prototype models were developed to try to create accounts of discrete concepts that could nevertheless explain people's sensitivity to the degree to which features correlate across the exemplars of a concept. In a prototype model, similarity can be viewed in geometric terms − the closer together items are in feature-space, the more similar they are. Thus more typical category members will be closer in space to the prototype, and less typical category members will more distant. Prototype models account well for findings relating to graded typicality, and offer a formal account of why new exemplars that are very prototypical are likely to be judged as being the better examples of a given category than items farther from the prototype.

However, these models fail to account for other findings in the literature. For example, prototype models do not store information about the frequency of specific exemplars, yet it is clear that people are sensitive to this information (Kruschke 1996). Moreover, the averaging process at the heart of prototype representations can yield anomalous results: If the members of a disjunctive category comprise either large black vertical lines or small white horizontal lines, then averaging across both dimensions produces a medium-sized grey diagonal line. This would fail to represent any of the relevant dimensions of the items associated with the concept appropriately, and measurements of similarity between this prototype and the actual set of members would not provide a good estimate of category membership.

Neither of these problems immediately falsify prototype models: there is no in-principle reason why exemplar frequency could not be incorporated into a prototype representation. Nor is there any reason why multiple prototypes could not be used to represent categories that are not linearly separable (although this might be hard to implement in a principled way). However, a more serious problem for prototype models is that they do not easily accommodate people's ability to recognize specific exemplars of concepts. For example, people asked to listen for the recurrence of a word in a lengthy, spoken wordlist do better when repeated words are presented by the same voice rather than a different one (Palmeri et al. 1993), suggesting that people store more auditory speech detail than linguists often suppose, and that models that store category summaries discard too much information about the speech signal to provide an adequate account of people's behavior.

3.3. Exemplar models

In an *exemplar model* (e.g, Nosofsky 1991, 1992) every example of a concept is stored in memory in all its detail. Novel items are classified by their similarity to previously learned exemplars, an category membership is determined by a weighted voting system in which a new item is assigned to the category for which the summed pairwise similarities are greatest (Kruschke 1992).

Interestingly, because of the way this voting process works, exemplar models are able to account for the typicality effects that led to the development of prototype models. This is because more typical exemplars, which, of course, lie near the center of the feature space of a category (the prototype), share more similarities with other exemplars than less typical items. Because the number of votes an item receives is a function of these similarities, a typical new item receives greater support than a less typical item.

Exemplar models have been tremendously influential, and yet what is perhaps their most important feature is usually least remarked upon in the literature: exemplar models do not contain, or even attempt to define, unitary representations for concepts. Instead, they typically contain a system of exemplars that is related to a system of labels, and a methodology for incorporating new items into this system and for dynamically generating labels for unlabeled items.

3.4. Systems models

Prototype models are often criticized for throwing away too much information, whereas exemplar models challenge our intuitions through the promiscuity of their assumptions about storage and processing. What is clear is that depending upon the context in which a concept is learned, or the goal of the learner, more or less information is discarded in learning. Moreover, the processes that appear to be involved in learning and storage inevitably result in encodings in which some stimulus dimensions are ignored in order to increase the discriminability of encoded items (Kruschke 2001; Ramscar et al. 2010b).

However, while it is unclear that a "pure" exemplar model even makes theoretical sense, simply because identifying exemplar is itself an act of classification at a different level of abstraction (Ramscar et al. 2010b), what is interesting about exemplar models is that they seek to capture people's behavior in tasks rather than seeking to define concepts: They treat categorization as in inherently systematic process relying on multiple concept representations, and yet they successfully account for many empirical phenomena (Nosofsky 1990).

The shift towards trying to model systematic behaviors rather than defining representations has led to models that employ multiple representations to find the middle ground between maximal abstraction (with minimal storage, e.g., prototypes) and minimal abstraction (with maximal storage, e.g., "pure" exemplar models). For example, the RATIONAL model of categorization (Anderson 1991) neither stores every exemplar nor does it rely entirely on prototypes. Instead, the model creates hybrid representations in which a new item may either be used to update an existing cluster of similar examples (as in a prototype model) or, if unique enough, it may initiate a new cluster. Which choice is made is a function of the probability that the new item belongs to an existing cluster. When this probability is below a given threshold, a new cluster is created. If above the threshold, the existing cluster that it is most similar is updated to reflect the new exemplar. RATIONAL is thus capable of acquiring clusters that function like rules, or sets of clusters that function like exemplars, depending on the categories being learned.

Other systems apply explicitly different mechanisms (rules initially; exemplars later) at different stages of concept learning (Johansen and Palmeri 2002), while in others (e.g., ATRIUM: Erickson and Kruschke 1998; COVIS: Ashby et al. 1998), the contributions of rule-based and exemplar learning are flexible, and depend more on the learning context, or the context in which categorization decisions are made.

Whereas most models seek to learn the representational system that best segregates a training set, a more recent clustering model, SUSTAIN (Love et al. 2004) was developed to account for the fact that people learn different information about items according to their context and goals. In unsupervised learning, when a learner has no specific goal in mind, SUSTAIN adds clusters much as RATIONAL would, in order to minimize classification error. However, if a learner is, say, inferring the properties of an item as part of a task, this goal can influence what is learned about the items. SUSTAIN is thus able to capture the differences in learning that occur in different task environments.

Depending on the conceptual structure being learned, and whether a goal is present or not, the structure of the clusters SUSTAIN learns for any given label can functionally resemble either rule-based, prototype or exemplar models. What is important to note is that the internal structure of these representations are highly sensitive to the context provided by the presence (or absence) of goals. Depending on context, different information will be represented in the clusters and different information discarded in learning or used in categorization. The success of the SUSTAIN model when it comes to fitting a wide range of behavioral phenomena suggests that people may learn different representations when learning concepts in inference and classification tasks and thus contributes further evidence that human category learning involves multiple processes, and that what is learned depends on a learners' goals and prior experience (Mack et al. 2013).

4. The neural bases of categorization

Results from cognitive neuroscience research support the findings reviewed so far in that they indicate there is no single neural circuit for concept learning (Seger and Miller 2010; Davis et al 2014), and suggest that categorization is best understood in relation to the overall architecture of the brain's perceptual- and motor- learning and memory systems. (Understanding the relevant neural architecture also requires an understanding of brain anatomy and physiology that few linguists currently have, so while this section may be challenging, we hope readers will appreciate its relevance to our understanding of concepts.)

4.1. Perceptual concepts

Numerous neural structures are involved in the discrimination of classes of visual stimuli, and even systems usually considered to be primarily engaged in perceptual processing exhibit evidence of tuning in response to categorization tasks: Different neuronal assemblies in the inferior temporal cortex (ITC) respond selectively to different category types, such as complex shapes (Logothetis and Scheinberg 1996) or faces (Kanwisher et al. 1997). Studies of trained monkeys have shown that individual neurons in the ITC show selectivity for, say, trees or fish, and these neurons are relatively insensitive to variance within these categories (Vogels 2008).

Human patients with impaired higher order memory systems (e.g., medial temporal lobe lesions) or Parkinson's disease (which impairs corticostriatal learning) retain their ability to implicitly learn prototype patterns presented in a series of random dot displays in a classification task (each pattern is labeled either "A" or "not A"; Bozoki et al. 2006). In contrast, imaging studies show that neuronal assemblies in the extrastriate visual cortex (roughly, Brodmann Area 18/19 or visual area V2) deactivate selectively when dot patterns that conform to a previously learned prototype are presented (Koenig et al. 2008), and patient studies have found that performance on this task is impaired in Alzheimer's disease, which often damages this area (Zaki et al. 2003).

However, although extrastriate assemblies appear to learn perceptual prototypes (whether this represents long-term potentiation or short-term adaptation is an open issue), learning and representing the range of discriminations manifest in visual categorization clearly involves a range of functional systems, with different brain regions involved in learning in different contexts (Seger and Miller 2010).

4.2. Higher-level concept learning

The prefrontal cortex (PFC) plays a key role in rule-based learning (Monchi et al. 2001), however, its contribution to the learning *process* is best described as supervising input to other learning systems in the striatum and MTL (Ramscar and Gitcho 2007; Thompson-Schill et al. 2009) both by maintaining representations of explicit goals and by allocating attentional resources (Miller and Cohen 2001).

This characterization is supported by imaging studies of rule-based learning (Konishi et al. 1999; Monchi et al. 2001; Smith et al. 1998) and behavioral experiments showing that tasks that disruption to working memory or attention (known PFC functions) drastically impairs performance on rule-based learning tasks (Waldron and Ashby 2001; Zeithamova and Maddox 2006). It is worth noting in this context that rule-based concept learning is very different from linguistic convention learning, where there is evidence that limiting PFC involvement actually benefits learning (Ramscar and Yarlett 2007; Ramscar et al. 2013a).

The actual learning systems most connected with the PFC, and which serve to discriminate the stimulus dimensions that encode concepts for long-term retention are located in the corticostriatal and medial temporal lobe regions (Seger and Miller 2010). The striatum (comprising the caudate, putamen, and nucleus accumbens, *Nacc*) implements an error-sensitive learning system that discriminatively strengthens and weakens associations between stimulus dimensions and behavioral responses and predicted outcomes in learning (Schultz 2006). In contrast to the perceptual regions described earlier, this system appears to learn the predictive, discriminatory codes that support future categorization behavior. However, the exact level at which striatal learning serves to encode concepts is open to question.

One reason why the role of striatal learning in categorization is hard to pin down is that the medial temporal lobe (MTL) also supports learning that is sensitive to prediction error (Davis et al. 2012a, 2012b; see Ramscar et al. 2010b for a tutorial), and also serves an encoding function. As with the striatal system, it is not clear *exactly* what part the MTL plays in concept learning or at what level of abstraction it encodes (indeed, it is not even clear whether this question is appropriate without reference to a learner's prior experience and goals). Some theories have proposed that the MTL learning system approximates an exemplar-model (Ashby and Maddox 2005); however evidence has been presented to suggest that the MTL stores representations of both rules (Nomura et al. 2007) and prototypes (Reber et al. 1998).[3]

---

[3] To confuse matters further, some researchers deny the MTL plays any role in category learning (Ashby et al 1998). However, this claim also conflicts with any broad definition of categoriza-

It is of course possible that the striatal system might learn one form of conceptual representation, and the MTL another (Bornstein and Daw 2012). However, in reality it is unlikely that the function of *either* system corresponds exactly to any of the models reviewed above. As Kruschke (2008: 269) observes in this regard, “[a] representational assumption for a model does not necessarily imply that the mind makes a formal representation[…] Only a formal model requires a formal description.” The brain’s actual representational formats should not be expected to correspond to the ones researchers use to model behavior. Similarly, it is unlikely that the functions of brain regions map neatly to functions posited by researchers, such that perception is a function of one system, and learning and categorization others. Indeed, Bussey and Saksida (2007) propose that brain function is based on a representational hierarchy of the levels of stimulus representations that systems process, rather than traditional cognitive functions like language or memory.

From this perspective, regional functions are differentiated by the levels of stimulus complexity they process (Cowell et al. 2010a, 2010b), and the striatum and MTL are not qualitatively different learning systems, but rather they learn and encode stimulus representations at different levels of abstraction (e.g., features, objects, contexts, etc.). Depending on the experience of the speaker and the context, learning might be focused in the MTL *or* in the striatum, and the degree each region is engaged in specific categorization behavior will depend on the experience of the individual and the behavioral context (Davis et al. 2012a, 2012b).

4.3. Categorization in the brain

Many neural systems contribute to the behaviors we call categorization: There is no categorization area, but rather, consistent with the predictions of systems models such as SUSTAIN, the degree to which brain regions engage in categorization depends on a task, its context and prior learning.

5. Concepts, contrasts and communication

From a linguistic perspective, the general lesson to be drawn from this review is that despite theorists’ intuitions about concepts as abstract mental tokens suitable for binding to phrase, word, or morpheme-sized phonetic patterns, this conception of concepts is not supported by research results. Rather, the literature shows:

1. Category assignments vary with context. An item can be an exemplar of one category in one context, and another category in another context.

---

tion, simply because perirhinal cortex – part of the MTL – appears critical to object recognition, and in particular, the discrimination of objects into old and new sets (Winters et al. 2004). Since old/new discrimination is thought of as an important categorization behavior, this debate helps illustrate how categorization is often poorly defined and circumscribed in the literature.

2. Even when people are learn concepts with clear and consistent definitions, the representations they acquire diverge from these definitions.
3. When people list the properties of natural concepts, they may produce convergent sets of features that characterize these concepts, but they generally do not adequately define or discriminate between concepts.
4. The tasks people perform when learning to apply concepts has a strong effect on the representations they acquire.
5. Depending on the task and context in which people perform the categorizations, it appears that a variety of brain regions are differentially engaged in the behaviors we call categorization.

It seems clear that in the course of learning the relationship between a set of items and a word, people do not abstract a discrete concept that specifies these relationships. Instead, categorization can engage almost any aspect of a person's knowledge, depending on their experience, the context and the task. These conclusions are supported by modeling efforts which show how conceptual behaviors are best accounted for by systems in which behavioral outcomes do not imply explicit knowledge representations (Baayen et al. 2013) and in which consistent conceptual behavior emerges without the use of the discrete representations our intuitive understanding of the word concept implies (Love et al. 2004). These conclusions are further supported by neuroscience findings revealing the equally varied and complex relationship between categorization behaviors and neural processes.

5.1. Learning, discrimination and language

If, as seems clear, concepts are not mental tokens, then explaining the role of words in communication is likely to depend on our understanding the processes governing word use. Providing a detailed account of the processes that support the learning and use of language is beyond the scope of this review (indeed, we do not pretend that we have a comprehensive account). However, one implication of these results is easy to state and is clearly important to understanding of language: it is clear to us that concept learning is a *discriminative process*. In explaining what this means, and why it is important, we will try to sketch out some of its implications for our understanding of human communication.

We noted above that the brain regions that support learning about lexical "concepts" implement *error-driven learning processes* (Schultz 2006). Most psychologists and linguists labor under the erroneous belief that learning is a combinatorial process in which correlations lead to simple associations forming between stimuli (Rescorla 1988). However, the error-driven learning processes that have been shown to govern what we call associative learning are *discriminative* (Ramscar et al. 2013c). That is, they reapportion attentional and representational resources to minimize future *predictive* uncertainty (Rescorla and Wagner 1972; Sutton and Barto 1998).

Importantly, although linguistic meaning is rarely couched in these terms, it is clear that that uncertainty reduction lies at the heart of communication: Virtually every linguistic act – even saying, "Hello!" – is intended to reduce a listener's uncertainty, whether about the world, about the thoughts and feelings of a speaker, or a speaker's sincerity, etc. Error-driven learning tunes the representation of relevant features of the environment by incrementally discriminating against uninformative cues (those that do not improve predictions) and reinforcing informative cues (those that do tend to support successful predictions) in response to events as they unfold (Ramscar et al. 2010b). It generates predictive representations that serve to minimize uncertainty in discriminating between sets of possible outcomes (i.e., about what a speaker might mean in a given context, or what verbal gesture might be uttered next).

These representations are formed by a process of learning to *ignore* − i.e., discriminate against cues that are not informative for the discrimination being learned. If learners learn the relationships between specific utterances in context and certain words by this process, it implies that they will learn about the contexts where the word is appropriately used and where it is not. The learning occurs as a result of speakers' attempts to predict the next segment, syllable or word that an interlocutor uses.

This in turn suggests that the relationship between categories (a domain with a huge array of dimensions) and labels (a domain with a relatively small number of phonetic dimensions) is subject to an important constraint. A naïve view of labels is that they serve to encode or otherwise map onto meanings, such that that they support the retrieval of underlying semantic categories. For example, it is usually supposed that a phonological pattern like *dog* serves as a pointer or a link to the concept of dog. However, the evidence we have reviewed indicates that learners acquire a variety of representations comprising information at numerous of levels of abstraction relating to a word like *dog*, and mapping a low-dimensional space of labels onto this set of high-dimensional representations in a determinate way is not possible (Ramscar et al. 2010b).

Linguists have assumed since Saussure that the relationship between words and meanings is bidirectional (suggested by the up and down arrows in Saussure's model for a *linguistic sign* connecting a graphic image of a tree − the meaning − with the orthographic Latin word *arbor*). The array of evidence indicating that conceptual learning processes are error-driven makes clear that this relationship must actually be unidirectional. Error-driven learning processes encourage the acquisition of representations in which word meanings − and other words in the context − are part of a high-dimensional predictive code that allows word identities to be discriminated and uncertainty about communicative intentions to be reduced. They are completely unsuited to acquiring representations in which words directly encode meanings such that Meanings predict Signals *and* Signals predict Meanings (Ramscar et al. 2010b).

It is interesting to consider representational proposals such as *frames*, *ICMs*, *image schemas*, and *domains* in this light. When these proposals are conceived of as theories of representation, it is assumed that at some level, something that resembles a frame (or ICM, image schema or domain) *corresponds* to a similar level of linguistic construction (or feature), such that at an appropriate level of granularity, the frame (ICM, image schema or domain) *represents* the meaning of a construction or feature (i.e., these structures are thought to facilitate the process of going from the signal to the meaning). We noted above that none of these proposals has been described with the specificity required to explain how this actually works, and it is likely that the problems of mapping spaces of different dimensionalities this would entail means that it is impossible to do so (Ramscar et al. 2010b). (Although these proposals fail to meet the criteria for theories of representation, they may still be useful phenomenological descriptions of some aspects of the knowledge encoded in discriminative linguistic and conceptual systems; Rosch 1978.)

5.2. The concept of information

The predictive, discriminative codes that error-driven learning processes generate share many properties with the codes that *information theory* specifies for artificial communication systems (Shannon 1948). It is thus worth highlighting that artificial information systems are not merely digital in the commonly understood sense that they make use of binary codes of ones and zeros, but also in the more interesting sense that in information theory, the "information" communicated in systems is broken down into a system of discrete, discriminable states that can be encoded by various combinations of ones and zeros.

Shannon (1948: 379) defines artificial communication as the process of:

> reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have meaning; that is they refer to or are correlated according to some system with certain physical or conceptual entities. *These semantic aspects of communication are irrelevant to the engineering problem.* [Our emphasis]

Artificial communication systems encode discriminable messages in a common *source code* (which defines a system for contrasting between messages) and a receiver makes use of this code in order to discriminate (select) the actual message that has been sent in a signal from other possible signals. There is no meaning in this signal itself, but rather in the context of the source code, the zeros and ones that each message comprises serve to incrementally reduce uncertainty about the actual message being received.

Although it may seem counterintuitive, the findings we have reviewed indicate that the way in which we learn new conceptual distinctions is best characterized as a process that increases either the number of words and phrases that our minds are able to discriminate or the range of contexts across which known words and phrases can be discriminated (Ramscar et al. 2010b). Thus, for example, Ramscar et al. (2013d) show that changes in people's ability to learn the paired association of arbitrary words across the lifespan can be accurately predicted if the process is modeled discriminatively. People's ability to learn frequently co-occurring pairs, like *lock* and *door* differs little with age, whereas word pairs like *jury* and *eagle* become increasingly difficult to learn. Because the latter co-occur extremely rarely, discriminative learning causes them to become *negatively* associated in a system predicting lexical events. However, since both words are relatively infrequent, it takes many years of language experience for speakers to learn to this relationship well enough to exploit the negative expectation of *eagle* given *jury* in everyday language use. Negative associations not only help explain why learning pairs like *jury* and *eagle* gets more difficult the older a speaker as compared to pairs like *lock* and *door*, they allow discriminative learning models (Rescorla and Wagner 1972) to quantitatively predict the changes in their learnability across the lifespan with remarkable accuracy (Ramscar et al. 2013d).

From a discriminative perspective, language learning can be characterized as acquiring and mastering a predictive code for a system of lexical and phrasal contrasts. Language production can then be viewed as the process of using this system to construct a message that best represents a speaker's intended meaning. A linguistic *signal* can be thought of as all of the conventional audible and visible behaviors of a speaking person (or, in written language, orthographic and other visual cues). Because the listener possesses a system of conventionalized knowledge relating semantic cues to signals that is similar to the one the speaker uses to construct her *message*, he is able to anticipate (that is, at least partially predict) the speaker's intended meaning by reconstructing the message from the signal itself. Other aspects of the message will be contained in the differences between what the speaker says and the learner predicts, and these differences will result in learning; an essential aspect of linguistic communication.

There is much consensus among cognitive linguists that intention reading − social prediction − is an important component of word learning (Tomasello 2003, 2008). The perspective we describe − which is a function of the way people learn to relate words to the world (Ramscar et al. 2013b) − simply extends intention reading to language processing more generally. Comprehension arises out of what listeners know − which enables them to predict a speaker − and what listeners *learn* from the speaker: identifying the words and constructions that a speaker actually says leads to learning about why a speaker made the choices they did.

Just as the source code lies at the heart of artificial communication systems, *linguistic codes* are the heart of language. The linguistic code is the entire conventionalized inventory of phones, words, idioms, expressions, collocations and constructions shared by a community of speakers and listeners that enable them to communicate. Importantly, rather than something that is explicitly encoded in the words of a message, meaning is merely implicit in the common linguistic code. The conventionalized, *systematic* relations that hold probabilistically between all the linguistic signals as well as between the signals and the world enable listeners to incrementally reduce uncertainty about the messages speakers send in context (Ramscar et al. 2010b; 2013c). In a linguistic signal, that is, in an utterance or piece of text, the occurrence of a word does not serve as a pointer to a concept, but rather in the context of the signal and the message, the word serves to reduce the listener's uncertainty about the speaker's intent. As is the case with an artificial communication system, the meaning is never "in" the signal. Instead the signal serves to reduce uncertainty in the listener's head about the actual intended meaning in the speaker's message (Ramscar and Baayen 2013).

5.3. Meaning and learning

Characterizing human communication in this way highlights a very obvious difference between human and artificial communication systems: Human communicators learn as they go, whereas most artificial systems don't. Thus whereas the goal of an artificial communication system is to send a signal that is predictable with $p = 1$, messages in human communication are rarely, if ever, intended to be perfectly predictable simply because they are *intended* to evoke or even highlight a listener's uncertainty about some aspect of what a speaker intends.

A number of researchers working on language have concluded that language understanding includes a process of making moment-to-moment predictions about what is coming next when listening to speech (Altmann and Mirkovic 2009; Kutas and Federmeier 2007). However this kind of moment-to-moment prediction has usually been seen as *assisting* the comprehension of linguistic signals that encode meanings in the traditional concept-by-concept ways that categorization research was expected to illuminate. Our review suggests this research can offer no such illumination simply because words do not *encode* meanings. Rather, because prediction drives learning, and because the function of learning is uncertainty reduction, prediction lies at the heart of linguistic communication. Seen from this perspective, moment-to-moment prediction in language does not merely help in the processing of language, but rather, because prediction drives learning, it is a critical part of the process that makes linguistic communication meaningful (Ramscar and Baayen 2013).

We end by acknowledging that although the view of communication we have sketched out manages to avoid many of the problems involved in appealing to illusory mechanisms like concepts, it paints a picture of language that is very different from traditional ideas, and that likely clashes with many researchers' beliefs about what language is. On the other hand, this perspective is still consistent with the impetus behind much work in cognitive linguistics in that its assumptions are shared with theories of learning and cognitive processing on multiple levels. And it is highly compatible with the findings of the research reviewed here.

6. Summary

Despite the way categorization researchers often describe their object of study, the detailed results of their work show that the representation of conceptual knowledge does not involve a neat story about inventories of individuated conceptual tokens. Rather, these results show that conceptual knowledge is as bound by context as language is itself (Malt 2013). Looking beyond naïve, intuitive conceptions of concepts, it is clear that cognitive linguists have much to learn from researchers' increasing understanding of the processes that give rise to systematic categorization behavior. We have sketched one way in which the insights that have arisen out of research into concepts and categories is likely to have an impact on our understanding of language. It will be fascinating to see what develops out of a richer synthesis of these lines of enquiry in the future.

7. References

Altmann, Gerry and Jelena Mirković
2009 Incrementality and prediction in human sentence processing. *Cognitive Science* 33: 1−27.
Anderson, John R.
1991 The adaptive nature of human categorization. *Psychological Review* 98: 409−429.
Arnon, Inbal and Michael Ramscar
2012 Granularity and the acquisition of grammatical gender: How order of acquisition affects what gets learned. *Cognition* 122(3): 292−305
Ashby, Gregory, Leola Alfonso-Reese, And Turken and Elliott Waldron
1998 A neuropsychological theory of multiple systems in category learning. *Psychological Review* 105(3): 442−481.

Ashby, Gregory, and Todd Maddox
2005 Human category learning. *Annual Review of Psychology* 56: 149−178.
Baayen, Harald, Peter Hendrix and Michael Ramscar
2013 Sidestepping the combinatorial explosion: Towards a processing model based on discriminative learning. *Language and Speech* 56(3): 329−347.
Barsalou, Lawrence
1983 Ad hoc categories. *Memory and Cognition* 11(3): 211−227.
Billman, Dorrit, and James Knutson
1996 Unsupervised concept learning and value systematicity: A complex whole aids learning the parts. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22: 458−475.
Bornstein, Aaron and Nathaniel D. Daw
2012 Dissociating hippocampal and striatal contributions to sequential prediction learning. *European Journal of Neuroscience* 35(7): 1011−1023.
Boroditsky, Lera and Michael Ramscar
2002 The roles of body and mind in abstract thought. *Psychological Science* 13(2): 185−189.
Bozokia, Andrea, Murray Grossman and Edward Smith
2006 Can patients with Alzheimer's disease learn a category implicitly? *Neuropsychologia* 44(5): 816−827.
Brooks, Lee, Geoffrey Norman and Scott Allen
1991 Role of specific similarity in a medical diagnostic task. *Journal of Experimental Psychology: General* 120(3): 278
Bruner, Jerome, Jacqueline Goodnow and George Austin
1956 *A Study of Thinking*. New York: John Wiley and Sons, Inc.
Bussey, Tim and Lisa Saksida
2007 Memory, perception, and the ventral visual-perirhinal-hippocampal stream: Thinking outside of the boxes. *Hippocampus* 17: 898−908.
Cienki, Alan
2007 Frames, Idealized Cognitive Models and domains. *The Oxford Handbook of Cognitive Linguistics*, 170−187. Oxford: Oxford University Press.
Collins, Allan and Ross Quillian
1969 Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior* 8(2): 240−247.
Cowell, Rosemary, Tim Bussey and Lisa Saksida
2010a Components of recognition memory: Dissociable cognitive processes or just differences in representational complexity? *Hippocampus* 20(11): 1245−262.
Cowell, Rosemary, Tim Bussey and Lisa Saksida
2010b Functional dissociations within the ventral object processing pathway: Cognitive modules or a hierarchical continuum? *Journal of Cognitive Neuroscience* 22: 2460−2479.
Crawford, Elizabeth, Janellen Huttenlocher, and Peder Hans Engebretson
2000 Category effects on estimates of stimuli: Perception or reconstruction? *Psychological Science* 11(4): 280−284.
Davis, Tyler, Bradley Love and Alison Preston
2012a Learning the exception to the rule: Model-based fMRI reveals specialized representations for surprising category members. *Cerebral Cortex* 22(2): 260−273.
Davis, Tyler, Bradley Love and Alison Preston
2012b Striatal and hippocampal entropy and recognition signals in category learning: Simultaneous processes revealed by model-based fMRI. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 38: 821−839.
Davis, Tyler, Gui Xue, Bradley Love, Alison Preston and Russell Poldrack
2014 Global neural pattern similarity as a common basis for categorization and recognition memory. *Journal of Neuroscience* 34(22): 7472−7484.

Erickson, Michael and John Kruschke
1998 Rules and exemplars in category learning. *Journal of Experimental Psychology: General* 127(2): 107−140.

Evans, Vyvyan
this volume 24. Time. Berlin/Boston: De Gruyter Mouton.

Fillmore Charles
1982 Frame semantics. In: Linguistic Society of Korea (ed.), *Linguistics in the Morning Calm*, 111−137. Seoul: Hanshin.

Frege, Gottlob
1892 Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik* 100: 25−50.

Gahl, Suzanne
2008 "Thyme" and "Time" are not homophones. The effect of lemma frequency on word durations in spontaneous speech. *Language* 84: 474−496

Goldstone, Robert and Alan Kersten
2003 Concepts and categories. In: A. F. Healy and R. W. Proctor (eds.), *Comprehensive Handbook of Psychology*, Volume 4:00 *Experimental Psychology*, 591−621. New York: Wiley.

Goodman, Nelson
1972 Seven strictures on similarity. In: N. Goodman (ed.), *Problems and Projects*. New York: The Bobbs-Merrill Co.

Hampton, James
1995 Testing the prototype theory of concepts. *Journal of Memory and Language* 34(5): 686−708.

Hull, Clarke
1920 Quantitative aspects of the evolution of concepts. *Psychological Monographs* XXVIII(1.123): 1−86.

Johansen, Mark and Thomas Palmeri
2002 Are there representational shifts during category learning? *Cognitive Psychology* 45(4): 482−553.

Johnson, Mark
1987 *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason*. Chicago: University of Chicago Press

Kanwisher, Nancy, Josh McDermott and Marvin Chun
1997 The fusiform face area: a module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience* 17(11): 4302−4311.

Koenig, Phyllis, Edward Smith, Vanessa Troiani, Chivon Anderson, Peachie Moore and Murray Grossman
2008 Medial temporal lobe involvement in an implicit memory task: Evidence of collaborating implicit and explicit memory systems from and Alzheimer's disease. *Cerebral Cortex* 18: 2831−2843.

Konishi, S., M. Kawazu, I. Uchida, H. Kikyo, I. Asakura, and Y. Miyashita
1999 Contribution of working memory to transient activation in human inferior prefrontal cortex during performance of the Wisconsin Card Sorting Test. *Cerebral Cortex* 9(7): 745−753.

Kruschke, John
1992 ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review* 99: 22−44.

Kruschke, John
1996 Base rates in category learning. *Journal of Experimental Psychology: Learning, Memory and Cognition* 22: 3−26

Kruschke, John
2001 Toward a unified model of attention in associative learning. *Journal of Mathematical Psychology* 45(6): 812−863.
Kruschke, John
2008 Models of categorization. *The Cambridge Handbook of Computational Psychology*, 267−301. Cambridge: Cambridge University Press.
Kutas, Martha, and Kara Federmeier
2007 Event-related brain potential (ERP) studies of sentence processing. In: G. Gaskell (ed.), *Oxford Handbook of Psycholinguistics*, 385−406. Oxford: Oxford University Press.
Labov, William
1973 The boundaries of words and their meanings. In: C.-J. N. Bailey and R. W. Shuy (eds.), *New Ways of Analyzing Variation in English*, 340−373. Washington, D.C.: Georgetown University Press.
Lakoff, George
1987 *Women, Fire, and Dangerous Things: What Categories Reveal About the Mind*. Chicago: University of Chicago Press.
Lakoff, George
1993 The contemporary theory of metaphor. *Metaphor and Thought* 2: 202−251.
Logothetis, Nikos and David Sheinberg
1996 Visual object recognition. *Annual Review of Neuroscience* 19(1): 577−621.
Love, Bradley
2002 Comparing supervised and unsupervised category learning. *Psychonomic Bulletin and Review* 9: 829−835.
Love, Bradley, Douglas Medin and Todd Gureckis
2004 SUSTAIN: A network model of category learning. *Psychological Review*111(2): 309−332.
Lusk, Christopher
2011 Conifer−angiosperm interactions: Physiological ecology and life history. *Smithsonian Contributions to Botany* 95: 158−164.
McCloskey, Michael and Sam Glucksberg
1978 Natural categories: Well defined or fuzzy sets? *Memory and Cognition* 6(4): 462−472.
Mack, Michael, Alison Preston and Bradley Love
2013 Decoding the brain's algorithm for categorization from its neural implementation. *Current Biology* 23: 2023−2027.
Malt, Barbara
2013 Context sensitivity and insensitivity in object naming. *Language and Cognition* 5: 81−97.
Malt, Barbara, Silvia Gennari and Mutsumi Imai
2010 Lexicalization patterns and the world-to-words mapping. In: B. Malt and P. Wolff (eds.), *Words and the Mind: How Words Encode Human Experience,* 29−57. Oxford: Oxford University Press
Medin, Douglas and Marguerite Schaffer
1978 Context theory of classification learning. *Psychological Review* 85(3): 207−238.
Miller, Earl and Jonathan Cohen
2001 An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience* 24(1): 167−202.
Monchi, Oury, Michael Petrides, Valentina Petre, Keith Worsley and Alain Dagher
2001 Wisconsin card sorting revisited: Distinct neural circuits participating in different stages of the task identified by event-related functional magnetic resonance imaging. *Journal of Neuroscience* 21: 7733−7741.

Nomura, Emi, Todd Maddox, Vincent Filoteo, David Ing, Darren Gitelman, Todd Parrish, Marchsel Mesulam and Paul Reber
2007 Neural correlates of rule-based and information-integration visual category learning. *Cerebral Cortex* 17(1): 37−43.
Nosofsky, Robert
1991 Tests of an exemplar model for relating perceptual classification and recognition memory. *Journal of Experimental Psychology: Human Perception and Performance* 17: 3−27
Nosofsky, Robert
1992 Similarity scaling and cognitive process models. *Annual Review of Psychology* 43(1): 25−53.
Palmeri, Thomas, Stephen Goldinger and David Pisoni
1993 Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 19(2): 309−328.
Port, Robert
2010 Language is a social institution: Why phonemes and words do not have explicit psychological form. E*cological Psychology* 22: 304−326.
Port, Robert and Adam Leary
2005 Against formal phonology. *Language* 81: 927−964
Posner, Michael, and Steven Keele
1967 Decay of visual information from a single letter. *Science* 158: 137−139.
Posner, Michael, and Steven Keele
1970 Retention of abstract ideas. *Journal of Experimental Psychology* 83(2): 304−308.
Quine, Willard Van Orman
1960 *Word and Object*. Cambridge: MIT Press.
Ramscar, Michael and Harald Baayen
2013 Production, comprehension and synthesis: A communicative perspective on language. *Frontiers in Language Sciences* 4: 233.
Ramscar, Michael, Melody Dye, Jessica Gustafson and Joseph Klein
2013a Dual routes to cognitive flexibility: Learning and response conflict resolution in the dimensional change card sort task. *Child Development* 84(4): 1308−1323.
Ramscar, Michael, Melody Dye and Joseph Klein
2013b Children value informativity over logic in word learning. *Psychological Science* 24(6): 1017−1023.
Ramscar, Michael, Melody Dye and Stewart McCauley
2013c Error and expectation in language learning: The curious absence of 'mouses' in adult speech. *Language* 89(4): 760−793.
Ramscar, Michael and Nichole Gitcho
2007 Developmental change and the nature of learning in childhood. *Trends In Cognitive Science* 11(7): 274−279.
Ramscar, Michael, and Ulrike Hahn
1998 What family resemblances are not: The continuing relevance of Wittgenstein to the study of concepts and categories. *Proceedings of the 20th Annual Conference of the Cognitive Science Society,* University of Wisconsin − Madison.
Ramscar, Michael, Peter Hendrix, Bradley Love and Harald Baayen
2013d 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 and Harald Baayen
2014 The myth of cognitive decline: Non-linear dynamics of lifelong learning. *Topics in Cognitive Science* 6: 5−42.
Ramscar, Michael, Teenie Matlock and Melody Dye
2010a Running down the clock: The role of expectation in our understanding of time and motion. *Language and Cognitive Processes* 25(5): 589−615.

Ramscar, Michael and Daniel Yarlett
2007 Linguistic self-correction in the absence of feedback: A new approach to the logical problem of language acquisition. *Cognitive Science* 31: 927−960.
Ramscar, Michael, Daniel Yarlett, Melody Dye, Katie Denny and Kirsten Thorpe
2010b The effects of feature-label-order and their implications for symbolic learning. *Cognitive Science* 34(6): 909−957.
Reber, Paul, Craig Stark and Larry Squire
1998 Contrasting cortical activity associated with category memory and recognition memory. *Learning and Memory* 5(6): 420−428.
Rescorla, Robert
1988 Pavlovian conditioning: It's not what you think it is. *American Psychologist* 43: 151−160.
Rescorla, Robert and Allan Wagner
1972 A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In: A. H. Black and W. F. Prokasy (eds.), *Classical Conditioning II: Current Research and Theory*, 64−99. New York: Crofts.
Rips, Lance, Edward Smith and Douglas Medin
2013 Concepts and categories: Memory, meaning, and metaphysics. In: K. J. Holyoak and R. G. Morrison (eds.), *The Oxford Handbook of Thinking and Reasoning*, 177−209. Oxford: Oxford University Press.
Rosch, Eleanor
1978 Principles of categorization. In: E. Rosch and B. B. Lloyd (eds.), *Cognition and Categorization*, 27−48. Hillsdale: Erlbaum.
Rosch, Eleanor, and Carolyn Mervis
1975 Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology* 7(4): 573−605.
Rosch, Eleanor, Carolyn Mervis, Wayne Gray, David Johnson and Penny Boyes-Braem
1976 Basic objects in natural categories. *Cognitive Psychology* 8(3): 382−439.
Sakamoto, Yasuaki and Bradley Love
2004 Schematic influences on category learning and recognition memory. *Journal of Experimental Psychology: General* 33: 534−553
Schultz, Wolfram
2006 Behavioral theories and the neurophysiology of reward. *Annual Review of Psychology* 57: 87−115.
Seger, Carol and Earl Miller
2010 Category learning in the brain. *Annual Review of Neuroscience* 33: 203−219.
Shannon, Claude
1948 A mathematical theory of communication. *Bell Systems Technical Journal* 27(3): 379−423.
Shepard, Roger, Carl Hovland and Herbert Jenkins
1961 Learning and memorization of classifications. *Psychological Monographs* 75: 13.
Smith, David and Paul Minda
2000 30 categorization results in search of a model. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26(1): 3−27.
Smith, Edward E., Andrea L. Patalano, and John Jonides
1998 Alternative strategies of categorization. *Cognition* 65(2): 167−196.
Smoke, Kenneth
1932 An objective study of concepts formation. *Psychological Monographs* XLII(191): 1−46.
Sutton, Richard, and Andrew Barto
1998 *Reinforcement Learning*. Cambridge: MIT Press
Taylor, John
2003 *Linguistic Categorization*. Oxford: Oxford University Press.

Tomasello, Michael
2003 *Constructing a Language: A Usage-Based Theory of Language Acquisition.* Cambridge: Harvard University Press.
Tomasello, Michael
2008 *Origins of Human Communication*. Cambridge: MIT Press.
Thompson-Schill, Sharon, Michael Ramscar and Evangelia Chrysikou
2009 Cognition without control: when a little frontal lobe goes a long way. *Current Directions in Psychological Science* 8(5): 259−263.
Vogels, Rufin
2008 Categorization of complex visual images by rhesus monkeys. Part 2:00 singlecell study. *European Journal of Neuroscience* 11(4): 1239−1255.
Waldron, Elliott, and Gregory Ashby
2001 The effects of concurrent task interference on category learning: Evidence for multiple category learning systems. *Psychonomic Bulletin and Review* 8(1): 168−176.
Winters Boyer, Suzanne Forwood, Rosemary Cowell, Lisa Saksida and Tim Bussey
2004 Double dissociation between the effects of peri-postrhinal cortex and hippocampal lesions on tests of object recognition and spatial memory: Heterogeneity of function within the temporal lobe. *Journal of Neuroscience* 24: 5901−5908.
Wittgenstein, Ludwig
1953 *Philosophical Investigations*. London: Blackwell.
Yamauchi, Takashi, Bradley Love, and Arthur Markman
2002 Learning non-linearly separable categories by inference and classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 28: 585−593
Zaki, Safa, Robert Nosofsky, Nenette Jessup and Frederick Unverzagt
2003 Categorization and recognition performance of a memory-impaired group. *Journal of the International Neuropsychological Society* 9(3): 394−406.
Zeithamova, Dagmar, and Todd Maddox
2006 Dual-task interference in perceptual category learning. *Memory and Cognition* 34(2): 387−398.
Zentall, Thomas, Edward Wasserman and Peter Urcuioli
2014 Associative concept learning in animals. *Journal of the Experimental Analysis of Behavior* 101(1): 130−151.

Michael Ramscar, Tuebingen (Germany)
Robert Port, Bloomington, (USA)

5. Abstraction, storage and naive discriminative learning