What makes humans different from other animals? Müller called language our rubicon, the unbreakable barrier between man and beast. Modern psychologists and linguists know a lot more about language than Müller did; nevertheless, many current hypotheses are not that different from Müller’s. Contemporary theorists tend to give the rubicon a name, whether it is recursion (Hauser, Chomsky, & Fitch, 2002), symbolic thought (Deacon, 1998), reference (Bloom, 2004), or some other. These names identify the construct within language that is claimed to be unique to humans. In Hauser et al.’s (2002) influential nomenclature, this construct constitutes the “faculty of language, narrow.” In other words, the rubicon.

From this perspective, the growing number of reports about language-trained animals is important and potentially problematic. If there really is a dividing line between human beings and other animals, then what are we to make of the growing menagerie that includes Kanzi the bonobo (Savage-Rumbaugh, 1994), Alex the parrot (Pepperberg & Pepperberg, 2009), various dolphins (Herman, Richards, & Wolz, 1984), or the pack of dogs including Rico (Kaminski, Call, & Fischer, 2004), Bailey (Griebel & Oller, 2012), and now Chaser (Pilley & Reid, 2011)? On such a view, these cases must either be miraculous discoveries or trivial mistakes.

In the case that an analyst can declare the behavior to be nonlinguistic, the language-trained animal is showing a trivial party trick—perhaps one that reflects the malleability of the animal and the skill of the trainer—but not a scientific discovery that constrains theories of language or human nature. On the other hand, if the analyst cannot reject the importance of the behavior, then it is a miraculous discovery, and a hypothesis about the nature of language must be rejected. This second outcome is not common.

An example of the failure of this binary strategy comes from the recursion debate. Hauser et al. (2002) proposed “recursion” as their rubicon. As a response, Gentner, Fenn, Margoliash, and Nusbaum (2006) claimed to show that starlings, an unrelated species, could discriminate center-embedded strings from tail-recursive strings, a key test case for one particular formal definition of recursion. But—assuming for a moment that we accept this evidence (cf. Corballis, 2007)—what is the inference we must make? On the rubicon viewpoint, we must reject that recursion of this sort is constitutive of the human-unique portions of the language faculty. Such a rejection would be a major discovery about the nature of human language.

CHASING THE RUBICON?

Chaser: Unlocking the Genius of the Dog Who Knows a Thousand Words

Where, then, is the difference between brute and man? What is it that man can do, and of which we find no signs, no rudiments, in the whole brute world? I answer without hesitation: The one great barrier between the brute and man is language. Man speaks, and no brute has ever uttered a word. Language is our Rubicon, and no brute will dare to cross it.

—F. Max Müller, Lectures on the Science of Language (1862), p. 324

REFERENCES
But the truth is likely to be both more banal and more complicated. Starlings are not particularly closely related to humans, and they certainly do not share any common ancestors that have the same pattern learning abilities. Whatever they are doing to discriminate strings may have little to do computationally with human linguistic abilities. Even as a case of convergent evolution, the claim is quite weak: Starlings do not show the same behavior as humans. They simply share—putatively—the cognitive ability to process strings in the same formal complexity class (cf. Perruchet & Rey, 2005). So it is hard to see why—absent an interpretation of the behavior as rubicon crossing—an interesting avian ability should reveal much about human language.

So why don’t we reject the rubicon? As an alternative, we could begin by acknowledging that human language is different in kind not because of a single computational feature but because it is the sum of many smaller divergences from animal communication systems that add up to something qualitatively unexpected (Pinker & Jackendoff, 2005). This multifactorial view is unsatisfying and underconstrained but more likely to be true. With this context, let us turn to Chaser, a fascinating case study that could nevertheless be more scientifically illuminating were it not quite so preoccupied with rubicon crossing.

**Chaser: Unlocking the Genius of the Dog Who Knows a Thousand Words**

Chaser is a border collie raised by John Pilley, a retired psychology professor at Wofford College. As Pilley’s (2013b) engaging narrative describes, by the time of writing Chaser had mastered the names of more than a thousand objects. Upon hearing one of these names, Chaser could fetch the matching object from a large comparison set with high accuracy and under a variety of stringent testing conditions: in public, with other speakers, and in controlled set-ups with no opportunity for social cueing. Ample video documentation attests to this achievement. Given that Chaser’s skill is consistent with—though quantitatively greater than—that of other language-trained dogs (Kaminski et al., 2004; Griebel & Oller, 2012), I do not see any reason to doubt the evidence.

It is hard to imagine anyone more well suited to training Chaser than Pilley, who retired after a career spent teaching and studying animal learning. From very early in Chaser’s life, Pilley devoted 4–5 hours a day to training word meanings through a series of carefully designed procedures. It is also abundantly clear that Pilley cares deeply about Chaser. This devotion—along with Chaser’s corresponding attachment—forms the motivational core that allows such a remarkable achievement.

Of course, one potential critique of this work has to do with the quantity of training Chaser received. In an earlier commentary on canine word learning, Bloom (2004) notes that it took the dog Rico 9 years to master 200 words, whereas human children learn thousands or tens of thousands during that time. I am not quite sure how useful such quantitative comparisons are; the necessary deemphasis on quantity alone for future studies is perhaps an important contribution of Pilley’s work with Chaser. Chaser’s 1,000 words represent a milestone that many children do not reach until some time in their third year (Mayor & Plunkett, 2011). Is that number different in kind from the 40,000 I am told by various online tests that I know?

Aside from this question of numerical line drawing (more rubicons to cross?), there is the question of whether these numbers reflect differences in mechanism or merely in experience and learning rate. In human language acquisition, there is an increasing appreciation of the immense amount of language input children receive; one back-of-the-envelope calculation suggests up to a million words a month in some cases, leading common words to be heard hundreds, thousands, or tens of thousands of times before they are first produced (Frank, Tenenbaum, & Gibson, 2013). And the amount of language children hear appears strongly related to their vocabulary size and facility with language (Hart & Risley, 1995; Hoff, 2003; Weisleder & Fernald, 2013). For children, and almost certainly for dogs as well, hearing more language—ideally in supportive, clear situations—leads to learning more words. The dogs probably learn slower, but that alone is not evidence for much of anything about the type of learning they are doing.

In addition to learning to recognize a large number of words, Chaser showed several other behaviors of interest. First, he was able to learn by exclusion, demonstrating success in a “mutual exclusivity” experiment of the type commonly used in studying word learning by human children (e.g., Bion, Borsovsky, & Fernald, 2013; Markman & Wachtel, 1988). Second, he was able to master taxonomic categories for objects, responding correctly to “toy,” “ball,” and “frisbee” as category labels that covered several objects that also had their own names. Finally, and perhaps most impressively, Chaser learned to respond appropriately to multword commands, distinguishing “paw lamb” (touch with your paw) from “nose...
lamb” (push with your nose) and even learning to respond correctly to commands such as “To Jeffrey . . . take decoy” (Pilley, 2013a). I will discuss learning by exclusion and multiword combinations in the next sections, as well as the nature of Chaser’s words.

Learning by Exclusion

The simplest “mutual exclusivity” experiment goes like this: A child is presented with two objects, one familiar and one novel, and is asked to give the experimenter the “dax” (a novel label that the child presumably has never heard before). Across dozens of studies, children as young as 14 months show evidence of this kind of behavior (Halberda, 2003; Markman, Wasow, & Hansen, 2003), although they do not tend to retain the link between the novel word and novel object until significantly later (Horst & Samuelson, 2008; Bion et al., 2013).

Although the phenomenon of mutual exclusivity is clear and highly replicable, the psychological sources of children’s behavior remain obscure. Some accounts have posited that children reason from basic principles about the structure of the lexicon (Markman, 1990) or about novel stimulus mappings (Mervis & Bertrand, 1994); others have proposed explanations of the same behavior in terms of pragmatic reasoning about language use in context (Clark, 1990). Still others, including me, have argued that this behavior could arise from general mechanisms of statistical inference (Regier, 2005; Frank, Goodman, & Tenenbaum, 2009). In principle, these explanations themselves are not mutually exclusive and could support the same behavior to different degrees at different times in development (Lewis & Frank, 2013).

Given this situation, I am not sure what we can infer from the fact of Chaser’s learning by exclusion. The evidence—as I see it, at least—does not warrant any reverse inference from an animal displaying learning by exclusion to a particular explanation of why the animal shows that behavior. There is one clue in Pilley’s account of the exclusivity experiments, however, tossed in apparently as a device to add narrative tension. Chaser showed extremely slow reaction times (more than a minute) during his first mutual exclusivity trials. These early delays, followed by faster performance in later trials, suggest that rather than exclusivity being the result of an innate inference, some kind of learning was almost certainly taking place.

In his demonstration with Chaser, Pilley attempts to address issues that were raised about previous demonstrations of learning by exclusion. I do not think he succeeds in this goal, though perhaps this has as much to do with the multiplicity of explanations for the mutual exclusivity behavior as it does with his experimental designs. Markman and Abelev (2004) note that Kaminski et al.’s (2004) demonstrations with Rico had no control condition testing whether Rico succeeded because of a baseline preference for novelty. Perhaps Rico would simply have retrieved a novel object even if there were no label, just because it was more attractive. Pilley repeated the same basic experiment that Kaminski et al. (2004) did, but he added a control condition in which Chaser—before any exclusion training—was asked to fetch familiar competitors in the presence of novel objects (which he did successfully). Pilley argues that this control condition shows that Chaser does not have a baseline novelty preference. It does not. It shows instead that Chaser’s baseline novelty preference is not stronger than his mapping of familiar labels. Thus, this control condition does not address the underlying issue raised by Markman and Abelev. Chaser could still have initially been guided by a weak novelty preference; in fact, novelty might well have been the critical cue in those initial very slow learning trials.

Chaser’s “Grammar”

Beyond the sheer quantity he aimed for, Pilley’s experiments on multiword utterances are perhaps the most novel aspect of his work with Chaser. As reported in Pilley (2013a), Chaser was able to comprehend sentences of the form “to frisbee take ball” (meaning “take the ball to the frisbee”). Chaser performed above chance (75% correct) in executing instructions of this type, using the names of objects that had not been specifically trained in this task.

A few aspects of this set of experiments stand out. First, the odd word order Pilley chooses is not an accident: It is optimized to Chaser’s cognition. In pilot work, Pilley attempted to use English word order but found that Chaser typically took the object named by the final word in his mouth, even though it was meant to be the location (e.g., “take the ball to the frisbee” means that you should pick up the ball, not the frisbee). This learnability finding suggests that Chaser’s cognitive limitations—in this case, perhaps a recency bias in memory—shape his ability to learn particular regularities. Suggestions about the effects of cognitive limitations on language structure have been made regarding human language as well (Christiansen & Chater, 2008), so it would be of substantial interest to understand what kinds of regularities Chaser can and

BOOK REVIEWS • 101
cannot learn (and perhaps how those relate to canine sequence processing abilities). Such studies might be more informative than the mere fact of success.

Second, although Chaser’s word order learning is a substantial accomplishment, it is worth noting that Pilley’s behaviorist vocabulary is not really equipped to describe the findings he gets. He writes that “as the farmer alters and labels the goals of the herding, such as herding the sheep toward or away from the farmer or to different locations, the dog infers that the name of the location specifies a given location to which the dog must herd the sheep without the farmer; thereafter, the dog needs only a single verbal command to take the sheep to the designated location” (Pilley, 2013a). This vocabulary does not have a way to describe how Chaser represents abstractions such as “location” or how he maps them to positions in a string of words. In that sense, it is strongly reminiscent of the circular accounts given in previous behaviorist analyses of language (Skinner, 1957).

Are Chaser’s Words Just Associations?
The question of whether human children’s words are “mere associations” has been the subject of a substantial debate both historically and in recent years (e.g., Sloutsky, Lo, & Fisher, 2001; Waxman & Gelman, 2009). The same thing can be asked about Chaser’s words. This debate often tends to be more circular than informative. We might instead ask two related questions: What is the nature of Chaser’s representation of words, and what are the computations (inferences) that he can use them in?

My own guess is that words are represented by humans as associations between linguistic forms and concepts. These associative links are long-term memories that obey the regularities governing such associations (Bahrick, Bahrick, Bahrick, & Bahrick, 1993; Frank et al., 2013). They may gain some privilege due to modality: Associations with words are highly salient, and words themselves are represented in much more detail than, say, arbitrary tones. But I would argue that what makes our language special is not the words themselves but the inferences we make about the use of words.

Human speakers use words to refer to the world and to communicate about both internal and external content. Furthermore, even in early childhood children seem to appreciate the nature of this usage: They are able to make a wide variety of powerful inferences from the fact and circumstances of language use (Baldwin, 1995; Vouloumanos, Onishi, & Pogue, 2012). This distinction between the long-term storage of words and the inferences that can be made from their use in context is critical to recent computational accounts of word learning (Frank et al., 2009; McMurray, Horst, & Samuelson, 2012). Distinguishing learning and inference also helps explain a number of empirical facts about children’s behavior in word learning experiments, such as the rich-get-richer nature of word learning or the previously mentioned ability to make mutual exclusivity inferences without later retaining the target word.

Seen from this perspective, Chaser might have similar representations of word meaning to human children, at least for the concrete nouns and noun category labels that he knows. What still bears exploration is the extent to which he can make inferences about the social role that labels play. I would love to see his performance in classic tasks by Baldwin (1991), in which cues to the speaker’s intention are pitted against salience and temporal contiguity. Or, further afield, he could even be tested in looking-time paradigms such as those of Vouloumanos et al. (2012) to test expectations about the communicative functions of language. Such explorations would shed light on the degree to which Chaser is able to use words flexibly as a marker of social communication.

Conclusions
Chaser is a touching book that reports on the product of an astonishing amount of sustained effort by Pilley. Certainly it feels like an endpoint that may not soon be repeated in canine language training. On the other hand, I am not sure what it tells us about human language. I say that reluctantly, because there is a lot to like about this work. But its primary contribution to the discussion of human language is to provide (further) evidence against a series of rubicons—learning by exclusion and multiword combination, to name two.

If we reject this kind of dichotomous thinking and instead consider the simultaneous contributions of many different abilities to human language, the interest in Chaser will be not that he can show particular behaviors but how he performs them. I would have loved to see experiments that show the circumstances under which Chaser cannot learn a word, to test hypotheses about the necessary ingredients for learning. How would he perform in a range of intention reading and social cue combination tasks (Baldwin, 1991; Hollich, Hirsh-Pasek, & Golinkoff, 2000)? Pilley also speculates that Chaser may have learned some mean-
meaningful words by overhearing them in conversation; it would be exciting to see Chaser’s performance in a learning-by-overhearing paradigm that directly tests this ability (Akhtar, 2005). Unfortunately the book and accompanying scientific papers do not say nearly as much about Chaser’s limitations as they do about his abilities.

Chaser is admirable attempt to jump across the rubicon. But if we abandon the notion of rivers to be crossed and move toward a multifactorial view of what makes human language unique, we need to acknowledge that what once seemed a river may be a wide, muddy swamp instead. Exploring this landscape may require more slogging and less jumping.

Michael C. Frank
Department of Psychology
Stanford University
450 Serra Mall (Jordan Hall)
Stanford, CA 94305
E-mail: mcfrank@stanford.edu

NOTES
Thanks to Dominic Massaro, Ellen Markman, and Mark Seidenberg for helpful comments.

REFERENCES
Frank, M. C., Tenenbaum, J. B., & Gibson, E. (2013). Learning and long-term retention of large-scale artificial languages. PLOS ONE, 8, e52500.


disPELLING MYTHS ABOUT THE RELATIONSHIP BETWEEN LANGUAGE AND THOUGHT

The Language Hoax: Why the World Looks the Same in Any Language

It seems like only yesterday that our inquiry into the literature came from mainstream journals like this one and its many descendents. Included in the required literature were only a few books. For the cognitive psychology revolution, the most popular were Ulric Neisser’s Cognitive Psychology and Don Norman’s Memory and Attention.

Today our curiosity is fed more by the mass media than mainstream journals. When is the last time you picked up The New York Times or The Wall Street Journal and did not see an article reporting some behavioral or neuroscience discovery? Not to mention the many TED Talks that are required viewing to stay informed of the exponential growth in so many areas of inquiry. Admittedly, our behavioral science has become increasingly interdisciplinary, but this does not account for the pervasiveness of many of the innovations and new discoveries in behavioral science in our popular media. A couple dozen articles selected from The New Yorker alone could bring an undergraduate up-to-date on many of the domains of inquiry in our psychological science. Popular books such as the one being reviewed here also give behavioral scientists much food for thought.

Book, Author, and Thesis

John Hamilton McWhorter’s little but thorough book, which he calls a manifesto, fits this mode of communication, although he is not a journalist by trade but is qualified as a bona fide linguist and certainly a brilliant and engaging one. McWhorter is an associate professor in the English and Comparative Literature department at Columbia University. He is a prolific author and political commentator on topics ranging from language to race relations. His specialty research studies how creole languages develop and their influences on language change. Although he is a senior fellow at the Manhattan Institute, a libertarian think tank, he characterizes himself as “a cranky liberal Democrat.” His TED Talk (McWhorter, 2013) is an engaging treatise demonstrating that texting is an