Variability and the emergence of abstraction from simple learning principles: Evidence from early word learning and reading.

Bob McMurray, Dept. of Psychology, University of Iowa

A critical problem in language acquisition is the ability to abstract away from surface details to arrive at more invariant or structural aspects of language. This is fundamental to all areas of language, from the acoustic signal, to morphology and syntax, and even to processes that are often seen as outside of the core linguistic capacity like reading or categorization. While the need for abstraction has often led researchers toward modular representations and language-specific learning mechanisms (Pinker & Ullman, 2002), connectionist models (McClelland & Rogers, 2003; Rumelhart & McClelland, 1986) offer a proof of concept that domain-general learning can also achieve such abstraction. A limitation of these models is that they often do not clarify the properties of the input or the learning mechanisms that enable such abstraction. In this talk, I present a series of studies that illustrate one principle which may guide children toward abstraction—variability in seemingly irrelevant aspects of the input.

The first case study examines the interaction of children's sensitivity to phonological detail in words and their ability to map words onto referents. Stager and Werker (1997) showed that 14 month old infants, who are adept at discriminating minimal pairs like *bih* and *dih,* struggle to map such words onto two different referents. Some explanations of this failure focus on the fact that word learning may have higher task demands than speech discrimination (Yoshida, Fennell, Swingley, & Werker, 2009). However, a series of studies from my lab (Galle, Apfelbaum, & McMurray, in press; Rost & McMurray, 2009, 2010) suggests that paradoxically 14 month olds can succeed in this task, if the stimuli are more variable. Intriguingly, variation in phonologically relevant cues (e.g., VOT for voicing) is not sufficient to drive learning. Rather variation must be targeted to factors that are irrelevant for discriminating phonemes like the talker voice or prosody. Computational modeling (Apfelbaum & McMurray, 2011) explains why this is the case. This model (Figure 1) uses simple associative mechanisms to link sound patterns and objects. Crucially if infants do not know to ignore talker voice, then in the single speaker versions of this task they will erroneously link a talker voice to both objects; however, with multi-talker variation, the associations between specific voices and the objects are spread out and can never control behavior. Thus, variability in talker voice allows infants to learn to ignore it, and to ultimately form a representation of the phonological word-form that abstracts away from this irrelevant aspect of the surface form.

The second case study illustrates how this principle can be applied to reading. In the early stages of readings children must learn Grapheme-Phoneme-Correspondence (GPC) regularities which map the spelling of a word onto its sound. For example, in a CVC word A is pronounced /**æ**/ (e.g., *MAT*), but in a CVCe word (e.g., *MATE*) it is pronounced /e**ɪ**/. As in many areas of language acquisition, there has been a vigorous debate over whether GPC regularities are learned and represented as a system of abstract rules, or are acquired via domain-general statistical learning mechanisms (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Plaut, McClelland, Seidenberg, & Patterson, 1996; Seidenberg & McClelland, 1989). However, this debate has not offered clarity into the conditions under which learners can acquire such abstractions. Here, in a study of 224 first graders learning a small set of GPC regularities involving vowels, we show that variability in the consonants (which are not relevant for these regularities) is essential for enhancing learning, and enabling the students to generalize to new words and new tasks (Figure 2, Apfelbaum, McMurray, & Hazeltine, in press). Once again, variability in irrelevant materials helps the learners achieve a more abstract representation. As in our work on word-learning, this process of abstraction may be a natural consequence of associative learning, suggesting a fairly domain-general way to achieve it.

Together, these illustrate a powerful principle that may help children achieve abstract representations of language using simple, domain general learning mechanisms.

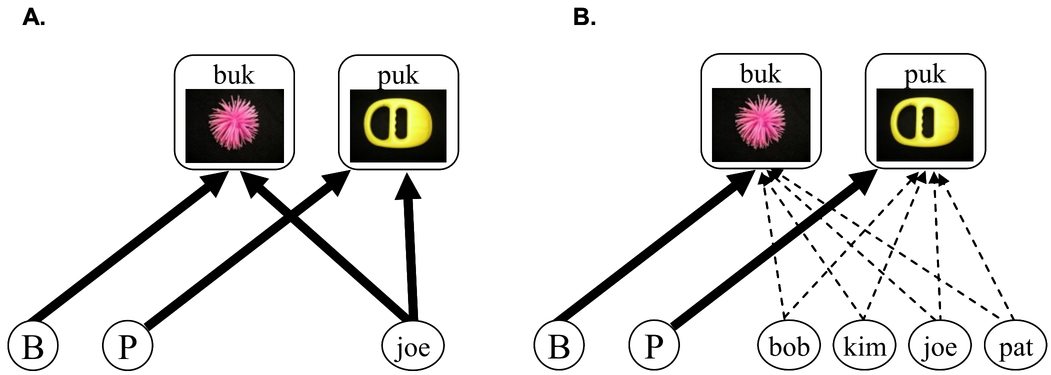

Figure 1: Associative model linking phonological features and words (bottom) to objects (top). A) When trained with a single talker (e.g., Joe), the model erroneously links the same talker to both objects. Later when the model hears either /b/ or /p/ in that voice at test, both objects are partially active. B) When trained with multiple talkers, no single talker ever forms a strong association with the object. As a result, its test behavior is driven more strongly by the relevant cues.

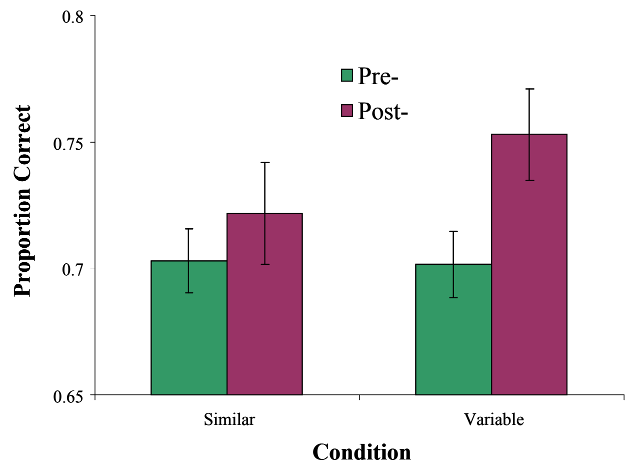

Figure 2: Performance on GPC regularities before and after training as a function of the variability or similarity of the consonant frames.

References

Apfelbaum, K., & McMurray, B. (2011). Using Variability to Guide Dimensional Weighting: Associative Mechanisms in Early Word Learning. *Cognitive Science, 35*(6), 1105-1138.

Apfelbaum, K., McMurray, B., & Hazeltine, R. E. (in press). Statistical learning in reading: Variability in irrelevant letters helps children learn phonics skills. *Developmental Psychology*.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*, 204-256.

Galle, M., Apfelbaum, K., & McMurray, B. (in press). Within-speaker variability benefits phonological word learning. *Language Learning and Development*.

McClelland, J. L., & Rogers, T. T. (2003). The Parallel Distributed Processing approach to semantic cognition. *Nature Reviews: Neuroscience, 4*(1).

Pinker, S., & Ullman, M. (2002). The past and future of the past tense. *Trends in Cognitive Sciences, 6*(11), 456-463.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103*(1), 56-115.

Rost, G. C., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental Science, 12*(2), 339-349.

Rost, G. C., & McMurray, B. (2010). Finding the signal by adding noise: The role of non-contrastive phonetic variability in early word learning. *Infancy, 15*(6), 608.

Rumelhart, D., & McClelland, J. L. (1986). On learning the past tense of English verbs. In J. L. McClelland, D. E. Rumelhart & t. P. R. Group (Eds.), *Parallel distributed processing: Explorations in the Microstructure of Cognition no. 2: Psychological and biological models*. Cambridge, MA: The MIT Press.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed developmental model of visual word recognition and naming. *Psychological Review, 96*, 523-568.

Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature, 338*, 381-382.

Yoshida, K. A., Fennell, C. T., Swingley, D., & Werker, J. F. (2009). Fourteen-month-old infants learn similar-sounding words. *Developmental Science, 12*(3), 412-418. doi: 10.1111/j.1467-7687.2008.00789.