Whorf for the 21st century: From interactive processing to linguistic relativity

Gary Lupyan (lupyan@wisc.edu)
Department of Psychology
University of Wisconsin-Madison
Madison, WI 53706

The crux of the Whorfian thesis is that our thought and behavior are influenced in deep ways by the language we use. In recent years we have seen a wave of rigorous and creative investigations of this thesis (Boroditsky, 2010; Wolff & Holmes, 2011 for reviews). Yet, many researchers remain highly skeptical of findings purporting to support Whorfian claims (Gleitman & Papafragou, 2005), and much confusion remains about how to integrate these findings into existing theories of cognition. A major barrier to understanding the degree to which various aspects of human cognition may be affected by speaking different languages is understanding the relationship between language—any language—and the rest of cognition. To remove this barrier we need to address a fundamental question: To what degree is normal human cognition actually language-augmented cognition? I will argue that a surprising variety of behavior previously assumed to be "nonverbal" shows signs of being influenced by linguistic factors and I will outline a theory of language-augmented thought that offers a mechanistic account of where we might expect to find effects of language on "nonverbal" cognition (Lupyan, 2012a, 2012b, for reviews).

One of the core features of language is using words to denote categories, e.g., using the word "dog" to refer to dogs. Words are commonly seen as a kind of "pointer" to concepts, the content of which is independent of language. In recent work, we have argued for an alternative: verbal labels do not simply point or refer to nonlinguistic concepts, but rather actively modulate conceptual representations that are brought online during "nonverbal" tasks. For example, Lupyan & Thompson-Schill (2012) showed that hearing referential labels such as "dog" consistently enhanced picture recognition compared to equally familiar, predictive, and unambiguous nonverbal cues such as a barking sound. This label advantage extended to newly learned labels and sounds. Despite participants' equivalent facility in learning what a novel object is called and what sound it makes, newly learned verbal labels were subsequently more effective in activating the concept than nonverbal sounds. In particular, hearing a label appeared to activate more category-typical information than hearing equally predictive nonverbal cues. This and related findings that verbal labels selectively activate category-typical features is hypothesized to underlie detrimental effects of labeling on visual memory such as the ability to remember not just that one saw a chair, but what kind of chair it was (Lupyan, 2008a).

As a further example of the kinds of powerful and surprising effects that category labels have on putatively nonverbal tasks, consider the following results (summarized in Lupyan, 2012a): When asked to draw a figure with three sides, all participants predictably drew triangles: 50% were isosceles/equilateral and 50% were parallel to the bottom of the page. When a separate group was asked to draw a "triangle," *91%* drew isosceles or equilateral triangles and *82%* drew triangles with bases parallel to the bottom of the page (the canonical horizontal orientation). These differences do not stem solely from pragmatics. In a speeded recognition task, participants were faster to verify isosceles than scalene triangles, and horizontally-oriented than oblique triangles, *but only* on trials on which they actually heard the word "triangle" and not on trials on which they viewed the same shapes after hearing "three-sided" (all factors within-subjects) . Finally, in an untimed visual-reasoning task, participants were asked to estimate the angle of the base of various three-sided polygons. These shapes were referred to as "triangle" or a "three-sided shape" (between subjects). As shown in **Fig. 1**, when the shape was referred to as a "triangle," its tilt was perceived as deviating more from the canonical (horizontal) as steeper than when the category name was omitted. On one interpretation, these results support the hypothesis that the representation activated by the word "triangle" are a better match to more "canonical" triangles than a formally equivalent cue. Despite denotative equivalence between "triangles" and "three-sided polygons" the category label "triangle" seems to reliably activate a more "canonical" triangle as measured by both explicit and implicit tasks—*prima facie* evidence of category labels augmenting underlying representations in systematic ways.

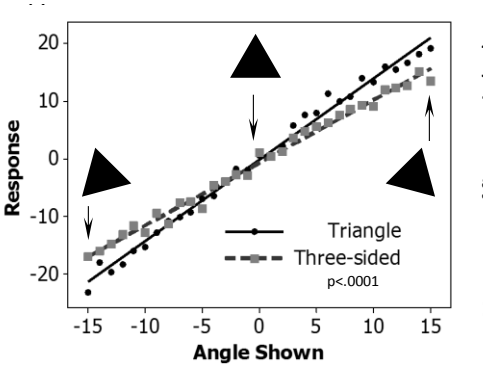

Figure 1: Left: Perceived orientation of shapes is systematically affected by whether they are called "triangles" or "three-sided". Right: Participants take longer to look at scalene (atypical) triangles when they hear the word "triangle."

Thus, referring to an object by its name appears to activate a *different* representation than when ostensibly the same concept is activated without using the name (Lupyan & Thompson-Schill, 2012; Lupyan, 2008b).

Why do labels have these effects? On the present account these effects are a product of (1) the association history of the discrete label with numerous category exemplars and (2) the *feedback* of the label on conceptual/perceptual representations. Under the influence of this feedback, the representations of various entities (objects, relations, etc.) become more categorical. This account can explain, for example, findings of pervasive effects of language on color perception (e.g., Regier & Kay, 2009; Lupyan, 2012a for discussion). Stated simply: the association of a label such as "green" with a range of colors means that when one sees a greenish color, the label is rapidly activated, *temporarily* warping the perceptual space. Viewing a green object becomes a hybrid visuo-linguistic experience.

To better understand this account, a simulation of how feedback label-feedback can augment conceptual and perceptual representations will be presented. **Fig 2A** shows a schematic of an interactive neural network trained on a bidirectional mapping between bit-vectors (representing feature-based object representations) and category labels (i.e., learning to label chairs as "chairs" and learning to activate a likely visual representation of a chair given the label). After training, when the network is presented with a perceptual input, the label becomes automatically activated, and then feeds back to affect representations as they unfold in time in the "perceptual" and "conceptual" layers. We can then examine what role the label is playing in the activation and maintenance of the representation of a particular category exemplar by directly perturbing the activation of the label or its feedback onto these layers. Feedback from labels (whether activated by the network on its own, or provided externally) provides much more categorical (clustered) representations **(Fig. 2b)**, leading to improved categorization—a prediction confirmed by overt categorization tasks (Lupyan, 2009; Lupyan, Rakison, & McClelland, 2007; Lupyan & Thompson-Schill, 2012). Additional evidence for verbal labels augmenting "nonverbal" representations comes from their apparent effects on basic visual perception. Visual representations activated by verbal means appear to be different—specifically, more categorical—than ostensibly the same representations activated by nonverbal means (Lupyan & Spivey, 2008, 2010; Lupyan, 2008b).

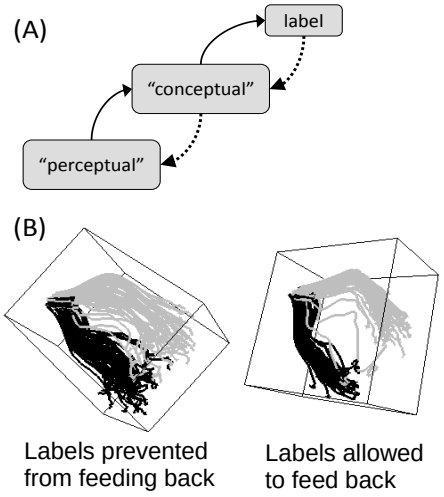

Figure 2. (A) Schematic of a neural network for studying the role of label feedback on object representations. (B) Activation dynamics in the "conceptual" layer. Each line represents an activation of a category exemplar over the course of a single trial. Color represents category membership.

Given that small linguistic manipulations affect how perceptual and conceptual information is brought online even within the same language community, we may expect that the substantial cross-linguistic differences in human languages should have substantially larger consequences on "thought," but there seems to be fewer such differences than expected. I will argue that this curious absence of evidence is due to a dichotomy made by researchers between verbal and nonverbal processes (e.g., "thinking for speaking") with the consequence that investigators may of linguistic relativity may have been looking in the wrong places.

References

Boroditsky, L. (2010). How the languages we speak shape the ways we think: the FAQs. In M. J. Spivey, M. Joanisse, & K. McRae (Eds.), *The Cambridge Handbook of Psycholinguistics* (p. forthcoming). Cambridge: Cambridge University Press.

Gleitman, L., & Papafragou, A. (2005). Language and thought. In K. Holyoak & B. Morrison (Eds.), *Cambridge Handbook of Thinking and Reasoning* (pp. 633–661). Cambridge: Cambridge University Press.

Lupyan, G. (2008a). From chair to "chair:" A representational shift account of object labeling effects on memory. *Journal of Experimental Psychology: General*, *137*(2), 348–369.

Lupyan, G. (2008b). The conceptual grouping effect: Categories matter (and named categories matter more). *Cognition*, *108*(2), 566–577.

Lupyan, G. (2009). Extracommunicative Functions of Language: Verbal Interference Causes Selective Categorization Impairments. *Psychonomic Bulletin & Review*, *16*(4), 711–718.

Lupyan, G. (2012a). Linguistically modulated perception and cognition: the label-feedback hypothesis. *Frontiers in Cognition*, *3*(54).

Lupyan, G. (2012b). What do words do? Towards a theory of language-augmented thought. In B. H. Ross (Ed.), *The Psychology of Learning and Motivation* (Vol. 57, pp. 255–297). Academic Press.

Lupyan, G., Rakison, D. H., & McClelland, J. L. (2007). Language is not just for talking: labels facilitate learning of novel categories. *Psychological Science*, *18*(12), 1077–1082.

Lupyan, G., & Spivey, M. J. (2008). Perceptual processing is facilitated by ascribing meaning to novel stimuli. *Current Biology*, *18*(10), R410–R412.

Lupyan, G., & Spivey, M. J. (2010). Redundant spoken labels facilitate perception of multiple items. *Attention, Perception, & Psychophysics*, *72*(8), 2236–2253.

Lupyan, G., & Thompson-Schill, S. L. (2012). The evocative power of words: Activation of concepts by verbal and nonverbal means. *Journal of Experimental Psychology-General*, *141*(1), 170–186. doi:10.1037/a0024904

Regier, T., & Kay, P. (2009). Language, thought, and color: Whorf was half right. *Trends in Cognitive Sciences*, *13*(10), 439–446. doi:10.1016/j.tics.2009.07.001

Wolff, P., & Holmes, K. (2011). Linguistic relativity. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(3), 253–265.