Current Directions in Psychological Science
2015, Vol. 24(4) 279–284

DOI: 10.1177/0963721415570732
cdps.sagepub.com

Words and the World: Predictive Coding and the Language-Perception-Cognition Interface

Gary Lupyan[1] and Andy Clark[2]
[1]Department of Psychology, University of Wisconsin–Madison and [2]School of Philosophy, Psychology, and Language Sciences, Edinburgh University

Abstract
Can what we know change what we see? Does language affect cognition and perception? The last few years have seen increased attention to these seemingly disparate questions, but with little theoretical advance. We argue that substantial clarity can be gained by considering these questions through the lens of *predictive processing*, a framework in which mental representations—from the perceptual to the cognitive—reflect an interplay between downward-flowing predictions and upward-flowing sensory signals. This framework provides a parsimonious account of how (and when) what we know ought to change what we see and helps us understand how a putatively high-level trait such as language can impact putatively low-level processes such as perception. Within this framework, language begins to take on a surprisingly central role in cognition by providing a uniquely focused and flexible means of constructing predictions against which sensory signals can be evaluated. Predictive processing thus provides a plausible mechanism for many of the reported effects of language on perception, thought, and action, and new insights on how and when speakers of different languages construct the same "reality" in alternate ways.

Keywords
perception, language, top-down effects, predictive coding, attention, language and thought

Across the cognitive sciences, a picture is emerging in which the brain is viewed as an engine of probabilistic prediction. On this view, every level of the hierarchically organized system that constitutes the brain works to predict the activity in the level below it (Fig. 1). A remarkable consequence of this arrangement is that seeking to reduce the overall prediction error produces representations at multiple levels of abstraction, flexibly incorporating whatever sources of knowledge help to reduce the overall prediction error. The higher-level (more abstract) representations formed in the goal of minimizing prediction error in the present enable better predictions in the future.[1]

The Predictive Brain

We begin with a brief outline of the predictive-processing framework and then apply it to two domains that, on the surface, seem to have little to do with each other but are unified under the new framework: the cognitive penetrability of perception, and effects of language on perception, action, and "thought" more broadly.

Consider the left-hand image in Figure 2—the so-called Cornsweet illusion. To most people, the central paired tiles appear to be very different shades of gray—an appearance that, as the right-hand picture reveals, is illusory. The illusion occurs because our visual experiences do not veridically reflect the current inputs but are informed by *priors* (prior beliefs, usually taking the form of nonconscious predictions or expectations) concerning the world. The relevant prior in this case is that surfaces tend to be equally reflectant rather than becoming gradually brighter or darker toward their edges. The image that produces the illusion displays a highly atypical combination of illuminance and reflectance properties, and the

Corresponding Author:
Gary Lupyan, Department of Psychology, University of Wisconsin–Madison, 1202 W. Johnson St., Madison, WI 53706
E-mail: lupyan@wisc.edu

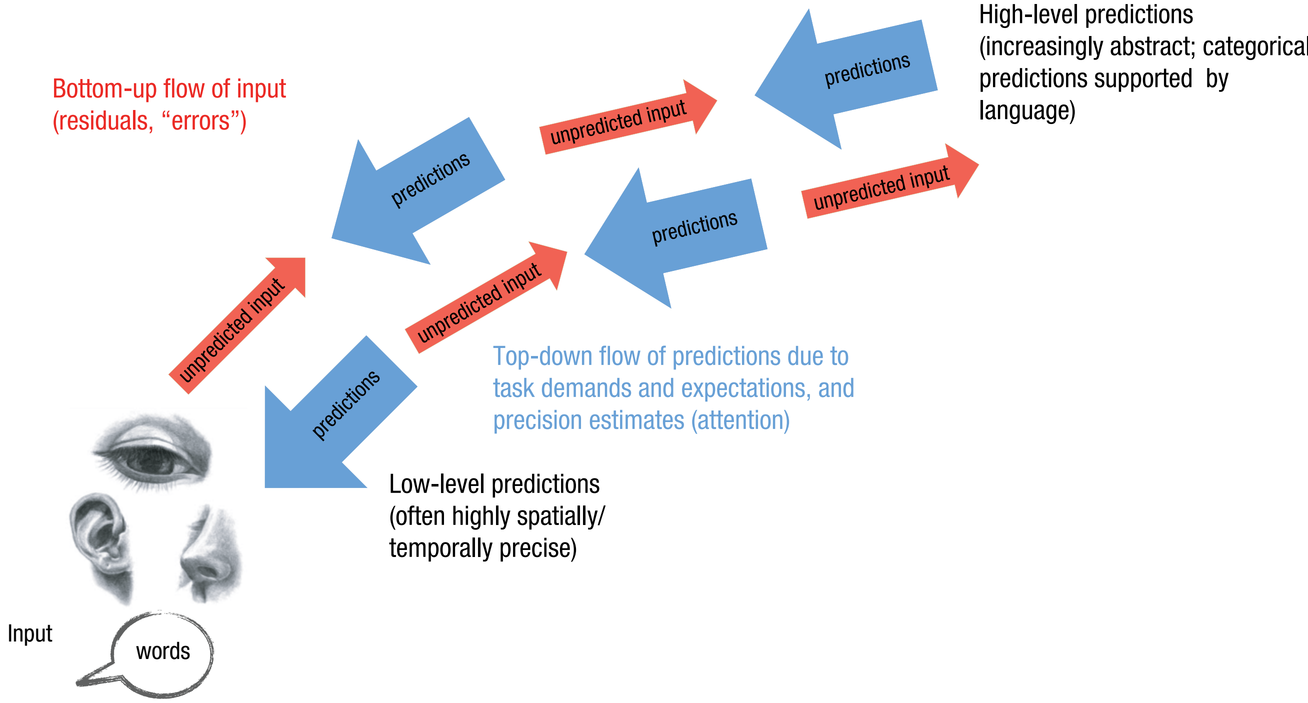

Fig. 1. A highly schematized view of the predictive-processing account of information transfer in the brain. Bottom-up inputs are processed in the context of priors (beliefs/hypotheses) from levels higher up in the hierarchy. The unpredicted parts of the input (errors) travel up the hierarchy, leading to the updating of subsequent predictions, and the cycle continues. The relative contribution of the bottom-up signal is determined by varying precision estimates. A highly variable or imprecise signal is given less weight.

brain uses what it has learned about typical patterns of illumination and reflectance to infer (falsely, in this case) that the two central tiles must be different shades of gray. In the world we actually live in, these particular prior beliefs or neural expectations are provably "Bayes optimal"—that is, they represent the globally best method for inferring the state of the world from the ambient sensory evidence (Brown & Friston, 2012).

This view of "perception-as-inference" originated with von Helmholtz (1867/2005) and has had many more recent champions, including Ulric Neisser and Richard Gregory. The brain, on these accounts, combines prior knowledge or expectations (including knowledge about the present context) with the incoming sensory evidence to yield a percept that reflects its best available hypothesis concerning the most probable state of the world. It is only in recent years, however, that these broad visions have been given effective computational flesh, shown to be (roughly speaking) neurally plausible, and seen to converge with compelling bodies of work in psychophysics and cognitive psychology showing that much of perception conforms to optimal (Bayesian) ways of combining sensory evidence with prior knowledge within the framework of predictive processing.

The predictive-processing framework (a term we use for models that implement hierarchical predictive coding; Clark, 2013; Friston, 2010; Hohwy, 2013) shares many features with earlier work on perception-as-inference and developments in connectionism/parallel distributed processing, such as McClelland and Rumelhart's interactive activation model (for some recent discussion, see McClelland, 2013; McClelland, Mirman, Bolger, & Khaitan, 2014). The predictive-processing framework adds an important emphasis upon hierarchical structure and the attempt to predict sensory signals (see also Hinton, 2007). A key emphasis of predictive-processing models is an asymmetry between the forward and backward flow of information: The forward flow computes residual errors, while the backward flow delivers predictions. Percepts emerge via a recurrent cascade of "top-down" predictions that involve expectations spanning multiple spatial and temporal scales. The downward predictions reflect what the system expects given what it already "knows" about the world and about the current context. These predictions are combined with incoming sensory data to arrive at progressively better guesses about the source of the signal (the world). Aspects of the input that are unexplained are sent forward as prediction-error signals that "carry the news" by pushing unexplained elements of the sensory signal upward so as to select new top-down hypotheses that are better able to accommodate the present sensory signal. This process

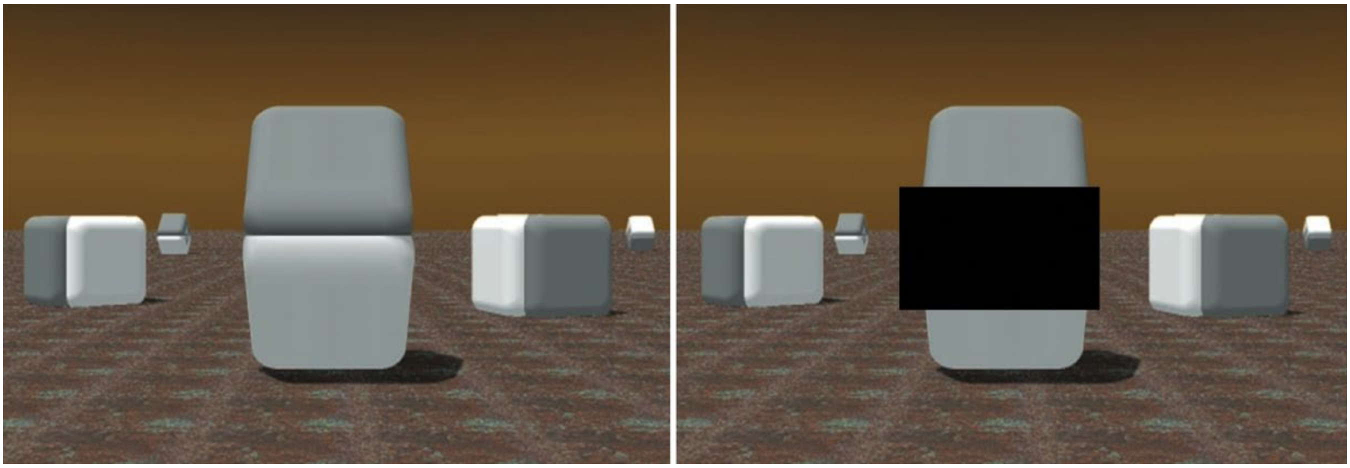

Fig. 2. The Cornsweet illusion. The image on the left depicts a typical Cornsweet illusion. The center of the two tiles comprising the central pairing appear to be different shades of gray. The image on the right reveals that they are, in fact, the same shade of gray. Reprinted from "An Empirical Explanation of the Cornsweet Effect," by D. Purves, A. Shimpi, and R. B. Lotto, 1999, *The Journal of Neuroscience, 19*(19), p. 8549. Copyright 1999 by the Society for Neuroscience. Reprinted with permission.

runs concurrently and continuously across multiple levels of a processing hierarchy.

While most of the predictions are unconscious, one can sometimes become aware of them when they are violated. For example, imagine drinking from a glass of what you think is orange juice only to realize on tasting it that it is actually milk. The difference between the taste of that milk when one expects it and when one expects orange juice instead is the orange-juice expectation made conscious (Lupyan, 2015, for discussion). Similarly, consider the experience of an unexpected omission, as when a musical note is missing from a familiar composition. Such omissions can be as perceptually striking and as salient as the most vibrant tone—an otherwise puzzling effect that is neatly explained by assuming that the construction of perceptual experience involves expectations based upon some kind of model of what is likely to occur.

The perceptual problems that confront us in daily life vary greatly in their demands. For some tasks, it is best to deploy large amounts of prior knowledge, while for others it may be better to let the world do as much of the driving as possible. Walking around our own house in the dark, it may be wise to let detailed top-down knowledge play a substantial role. Driving fast along an unfamiliar winding mountain road, we need to let sensory input take the lead. How is a probabilistic-prediction machine to cope? It copes by continuously estimating and re-estimating its own *sensory uncertainty*, assigning more or less weight to top-down expectations versus bottom-up inputs in the service of minimizing overall prediction error. Within this framework, estimations of sensory uncertainty modify the impact of prediction-error signals at each level of processing according to their estimated precision, which in this case is the brain's best guess at their certainty or reliability (*inverse variance*, for the statistically savvy). Variable precision weighting is thus a mechanism for tuning the extent to which input is modulated by top-down predictions. As we shall see, this mechanism also provides a way for language to serve as a superbly flexible tool for tuning sensory processing.

Perception as a Predictive and Penetrable Process

Viewing perception as a predictive process helps to resolve a long-standing argument concerning whether perception is "penetrated" by knowledge (Pylyshyn, 1999). Within the predictive-processing framework, perception is expected to be penetrable to the extent that such penetration minimizes overall (long-term) prediction error (Lupyan, 2015). If information from prior experience, expectations, knowledge, beliefs, and so forth lowers overall prediction error, then this information will be used to guide perceptual processing (we reiterate that this process is not a "decision" made by the organism but the consequence of minimization of the prediction error). In some cases, this penetration changes what we consciously experience (e.g., the lightness of the tiles in Fig. 2). In other cases, the conflict between bottom-up inputs and top-down predictions can be resolved at a higher level. For example, in Figure 3, the meaning of the central image changes depending on its context, but what we literally see is (relatively) unaffected (at least when we are free to examine the image at our leisure).

Fig. 3. Another example in which the prior local contextual cues set up expectations. When the central character(s) are processed in the letter context, the B hypothesis makes the raw visual data most probable. When processed in the number context, the 13 hypothesis makes the very same raw visual data most probable. Unlike the Cornsweet illusion, the central image remains visually ambiguous even as its meaning is disambiguated because the top-down prediction can be integrated with the bottom-up signal at a relatively higher level.

There is, of course, no gatekeeper deciding the extent to which a cognitive state should penetrate perception. In contexts where altering the activation patterns at lower levels of processing minimizes overall prediction error, we should find that what we know changes what we see. In other situations, the conflicts between predictions and inputs are resolved at higher levels that are sometimes referred to as *decisional* or *post-perceptual* stages. These are cases in which our perceptual phenomenology is relatively unaffected by knowledge and expectations. The predictive-processing framework offers a precise way to strike this balance according to the estimated reliability of the prediction-error signal at different levels of processing.

This formulation in terms of predictive processing helps to resolve two persistent confusions. First, it is commonly argued that if what we knew changed what we saw, then knowing that the two tiles in Figure 2 are actually the same lightness ought to cause us to see them that way (e.g., Pylyshyn, 1999). The problem is that discounting the input in this way is incompatible with long-term error reduction. If the illusory percept offers the best prediction in the majority of situations, then, in the long term, the illusion is Bayes optimal. Simply letting a belief override a bottom-up input will, in many cases, result in very high prediction error; the input and the higher-level belief need to be weighted according to their respective likelihoods. Second, critics of cognitive penetrability contend that many demonstrations of effects of beliefs, knowledge, and expectations on perception are merely attentional, such that knowledge can affect what one attends to but not how the attended inputs are subsequently processed (see Lupyan, 2015, for review). In contrast, within the predictive-processing framework, attention is not something one "focuses" or "deploys" (Anderson, 2011). Rather, it is the mechanism of variable precision weighting itself. When one "attends" to something, small deviations from expectations are weighed more than when one is not attending to it (Den Ouden, Kok, & de Lange, 2012; Feldman & Friston, 2010). As a result, the neural representations of an object that is being attended (because it is task relevant) are measurably different than those of the same object when it is not attended (e.g., Çukur, Nishimoto, Huth, & Gallant, 2013).

Predictive Processing and the Relationship Among Language, Perception, and "Thought"

A commonly held view is that the sole function of language is to communicate our thoughts. On this view, words and larger linguistic constructions latch onto preexisting concepts, enabling highly flexible communication, but do not alter the workings of "nonverbal" systems involved in, for instance, categorization, memory, and perception (Pinker, 1994; Snedeker & Gleitman, 2004). A corollary of this view is that although different languages provide their speakers with different ways of *talking about* things (Malt et al., 2015), these differences have nothing to with how we *think about* or *perceive* things (Gleitman & Papafragou, 2005).

A flurry of findings from cognitive and developmental psychology, however, argue for a much more transformative role of language both in higher-level cognition and in basic perception (for reviews, see Boroditsky, 2010; Casasanto, 2008; Lupyan, 2012). Language not only functions as a means of communicating our thoughts but plays an active role in shaping them. Rather than passively reflecting the joints of nature, words and larger constructions help carve joints *into* nature. For example, controlled studies of the famous "Eskimo words for snow" thought experiment show that, under certain conditions, named categories are easier to learn than equally salient unnamed categories (Lupyan, Rakison, & McClelland, 2007). Once learned, verbal labels continue to be uniquely effective in activating conceptual content (Boutonnet & Lupyan, 2015; Lupyan & Thompson-Schill, 2012). Considering the relationship between language and thought within the framework of predictive processing allows us to go beyond these individual observations toward a fuller, more unifying account.

Words as artificial contexts

We take for granted that we can change people's behavior using language. It is tempting to think of this ability in terms of language activating a repository of stored knowledge (a mental lexicon). A very different perspective is that language input directly affects mental states just like other perceptual inputs (Elman, 2009). For example, in a study by Çukur and colleagues (2013), participants undergoing fMRI watched movie clips passively or while monitoring for humans or for vehicles. The verbal prompts to attend to one category or another shifted neural representations throughout the brain (including primary visual cortex), such that a verbal prompt to attend to vehicles expanded the neural representations of vehicles and semantically related entities while collapsing semantically distant categories.[2]

Importantly, words and larger verbal constructions are *special* kinds of perceptual inputs. While perceptual experiences of, for example, vehicles are always experiences of *specific* vehicles, the word "vehicle" is *categorical*.[3] Verbal cues (even if self-generated) can therefore act as highly flexible contexts (sets of priors) within which an organism can appropriately weight the incoming input, possibly through a broad, fast retuning of the organism's entire semantic network, of the sort shown by Çukur et al. (2013). If this is true, we might expect that simply hearing a word can lead the visual system to generate a predictive signal helping to process an input that is otherwise too weak or noisy. Indeed, in a recent study, simply hearing a word boosted otherwise invisible images of items matching the named category into awareness (Lupyan & Ward, 2013).

Viewed from the perspective of predictive processing, language directed at others and at oneself (e.g., in verbal rehearsal and other forms of self-directed speech) becomes a powerful tool for manipulating thought and reasoning. Words (and larger verbal constructions) become not simply ways to communicate our preexisting thoughts but highly flexible (and metabolically cheap) sources of priors throughout the neural hierarchy. This is accomplished both through flexible modification of what top-down information is brought to bear and by selectively influencing the precision weighting of prediction error, thereby influencing how much top-down information influences specific lower-level level processes. This enables language to act as an "artificial context," helping constrain what representations are recruited and what impact they have on reasoning and inference.

By serving as a domain-general prior-setting tool, words and larger constructions thus afford a kind of flexible "programming language" for the mind (Lupyan & Bergen, in press), potentially providing a huge boost to intelligence.[4] As expected on such a position, language is persistently linked to an enormous range of behaviors. To take one example, vocabulary size and other verbal measures are surprisingly good predictors of performance on "nonverbal" intelligence tests such as Raven's Progressive Matrices (e.g., Cunningham & Stanovich, 1997), while linguistic impairments are linked to marked deficits (Baldo, Bunge, Wilson, & Dronkers, 2010).

In sum, we propose that the learning of language may create a potent means of biasing the recruitment of prior knowledge and of artificially manipulating, at any level of processing, the weightings that determine the relative influence of different top-down expectations and incoming sensory signals. These manipulations could selectively enhance or mute the influence of any aspect, however subtle or complex, of our own or another agent's world model. Exposure to language (whether shared or self-produced) thus becomes a potent and fundamentally unified means of exploring and exploiting the full potential of our own acquired knowledge about the world—a kind of artificial "second system" enabling us to take full advantage of our own knowledge as well as the knowledge of others.

Recommended Reading

Clark, A. (2013). (See References). Explores predictive processing as a unifying framework for understanding neural processing.

Frith, C. D., & Frith, U. (2012). Mechanisms of social cognition. *Annual Review of Psychology*, *63*, 287–313. doi:10.1146/annurev-psych-120710-100449. Applies principles of predictive coding to social cognition and mentalizing.

Feldman, H., & Friston, K. J. (2010). (See References). Summarizes computational accounts of hierarchical predictive coding and its relationship to competing theories.

Friston, K. J. (2010). (See References). Like Feldman and Friston (2010), summarizes computational accounts of hierarchical predictive coding and its relationship to competing theories.

Hohwy, J. (2013). (See References). Like Clark (2013), explores predictive processing as a unifying framework for understanding neural processing.

Lupyan, G. (2015). (See References). Reviews the evidence for and against cognitive penetration of perception, focusing on predictive coding.

Lupyan, G. (2012). (See References). Reviews evidence of effects of verbal labels on cognition and perception within a theory of "language-augmented thought."

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Notes

1. To be clear, the brain is no more "trying" to predict than a gas at a higher pressure tries to diffuse to a lower pressure. Living systems can temporarily resist such increases in entropy by avoiding some environments and altering others. Organisms that can predict their own sensory inputs at multiple spatial and temporal scales are well placed to do this, thus maintaining themselves within their species-specific window of viability (Friston & Stephan, 2007).

2. Such findings are consistent with claims from the field of embodied cognition that words activate neural patterns overlapping with those activated by nonverbal sensory inputs (Lupyan & Bergen, in press, for discussion).

3. This claim is not limited to superordinate terms such as "vehicle" and holds at any level of abstraction. Experiences with a

dog, the color red, an instance of on-ness, or your brother Bob are all *particulars*. The corresponding terms ("dog," "red," "on," "Bob") abstract over the particulars in a way that perception cannot.

4. The role of language in intelligence is taken for granted in much of classical and contemporary philosophy of mind, though without much elaboration of mechanism. In contrast, as noted above, major strands of contemporary cognitive science and developmental psychology dismiss language as being purely a tool for communication.

References

Anderson, B. (2011). There is no such thing as attention. *Frontiers in Theoretical and Philosophical Psychology*, *2*, Article 246. Retrieved from http://journal.frontiersin.org/article/10.3389/fpsyg.2011.00246/full

Baldo, J. V., Bunge, S. A., Wilson, S. M., & Dronkers, N. F. (2010). Is relational reasoning dependent on language? A voxel-based lesion symptom mapping study. *Brain and Language*, *113*, 59–64. doi:10.1016/j.bandl.2010.01.004

Boroditsky, L. (2010). How the languages we speak shape the ways we think: The FAQs. In M. J. Spivey, K. McRae, & M. Joanisse (Eds.), *The Cambridge handbook of psycholinguistics* (pp. 615–632). New York, NY: Cambridge University Press.

Boutonnet, B., & Lupyan, G. (2015). Words jump-start vision: A label advantage in object recognition. *The Journal of Neuroscience*, *32*, 9329-9335.

Brown, H., & Friston, K. J. (2012). Free-energy and illusions: The Cornsweet effect. *Frontiers in Psychology*, *3*, Article 43. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3289982/

Casasanto, D. (2008). Who's afraid of the big bad Whorf? Crosslinguistic differences in temporal language and thought. *Language Learning*, *58*, 63–79.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral & Brain Sciences*, *36*, 181–204.

Çukur, T., Nishimoto, S., Huth, A. G., & Gallant, J. L. (2013). Attention during natural vision warps semantic representation across the human brain. *Nature Neuroscience*, *16*, 763–770. doi:10.1038/nn.3381

Cunningham, A. E., & Stanovich, K. E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology*, *33*, 934–945.

Den Ouden, H. E. M., Kok, P., & de Lange, F. P. (2012). How prediction errors shape perception, attention, and motivation. *Frontiers in Psychology*, *3*, Article 548. Retrieved from http://journal.frontiersin.org/article/10.3389/fpsyg.2012.00548/full

Elman, J. L. (2009). On the meaning of words and dinosaur bones: Lexical knowledge without a lexicon. *Cognitive Science*, *33*, 547–582. doi:10.1111/j.1551-6709.2009.01023.x

Feldman, H., & Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience*, *4*, Article 215. Retrieved from http://journal.frontiersin.org/article/10.3389/fnhum.2010.00215/full

Friston, K. J. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*, 127–138. doi:10.1038/nrn2787

Friston, K. J., & Stephan, K. E. (2007). Free-energy and the brain. *Synthese*, *159*, 417–458.

Gleitman, L., & Papafragou, A. (2005). Language and thought. In K. Holyoak & B. Morrison (Eds.), *Cambridge handbook of thinking and reasoning* (pp. 633–661). New York, NY: Cambridge University Press.

Hinton, G. E. (2007). Learning multiple layers of representation. *Trends in Cognitive Sciences*, *11*, 428–434. doi:10.1016/j.tics.2007.09.004

Hohwy, J. (2013). *The predictive mind*. New York, NY: Oxford University Press.

Lupyan, G. (2012). 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). Waltham, MA: Academic Press. Retrieved from http://www.sciencedirect.com/science/article/pii/B9780123942937000078

Lupyan, G. (2015). Cognitive penetrability of perception in the age of prediction: Predictive systems are penetrable systems. *Review of Philosophy and Psychology*. Advance online publication. doi:10.1007/s13164-015-0253-4

Lupyan, G., & Bergen, B. (in press). How language programs the mind. *Topics in Cognitive Science*.

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

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*, 170–186. doi:10.1037/a0024904

Lupyan, G., & Ward, E. J. (2013). Language can boost otherwise unseen objects into visual awareness. *Proceedings of the National Academy of Sciences, USA*, *110*, 14196–14201. doi:10.1073/pnas.1303312110

Malt, B. C., Gennari, S. P., Imai, M., Ameel, E., Saji, N., & Majid, A. (2015). Where are the concepts? What words can and can't reveal. In E. Margolis & S. Laurence (Eds.), *The conceptual mind: New directions in the study of concepts* (pp. 291–326). Cambridge, MA: MIT Press.

McClelland, J. L. (2013). Integrating probabilistic models of perception and interactive neural networks: A historical and tutorial review. *Frontiers in Psychology*, *4*, Article 503. Retrieved from http://journal.frontiersin.org/article/10.3389/fpsyg.2013.00503/full

McClelland, J. L., Mirman, D., Bolger, D. J., & Khaitan, P. (2014). Interactive activation and mutual constraint satisfaction in perception and cognition. *Cognitive Science*, *38*, 1139–1189. doi:10.1111/cogs.12146

Pinker, S. (1994). *The language instinct*. New York, NY: Harper Collins.

Pylyshyn, Z. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral & Brain Sciences*, *22*, 341–365.

Snedeker, J., & Gleitman, L. (2004). Why is it hard to label our concepts? In D. G. Hall & S. R. Waxman (Eds.), *Weaving a lexicon* (Illustrated ed., pp. 257–294). Cambridge, MA: MIT Press.

von Helmholtz, H. von. (2005). *Treatise on physiological optics* (Vol. 3, J. P. C. Southall, Ed.). New York, NY: Dover. Original translation published 1924, original work published 1867