Language augmented prediction

Gary Lupyan*

University of Wisconsin Madison, Madison, WI, USA
*Correspondence: lupyan@wisc.edu

Edited by:
Shimon Edelman, Cornell University, USA

Reviewed by:
Axel Cleeremans, Université Libre de Bruxelles, Belgium

A commentary on

Whatever next? Predictive brains, situated agents, and the future of cognitive science
by Clark, A. (in press). Behav. Brain Sci.

Clark's (in press) article makes a strong argument that prediction or reduction of "surprisal" comprises a synthesizing principle in understanding neural mechanisms. But if brains – all brains – are "essentially prediction machines," how do we account for the apparently qualitative differences between humans and non-human animals in the ability to inspect and reflect on one's mental states, and to effectively foresee the consequences of various actions? For example, Spelke (2003) points out that although all animals find and recognize food, only humans developed the art and science of cooking. Although all animals have to understand (and predict!) the material world, only humans systematize their knowledge as science (p. 277). But we do not need to go into something as complex as formalized science to see the wide gap between human and non-human minds.

Imagine the simple task of pointing to a red box to get a reward, while ignoring the blue box. We can think of success as the mapping between the sensory input and the motor output that minimizes surprisal. Many animals can succeed on this task after being trained – their behavior nudged gradually by rewards until the generated predictions match the contingencies of the task. In contrast, humans can succeed without any training at all, simply by being told what to do![1] We often take this ability for granted, but without it, all human learning would require direct experience with the domain (e.g., see Carvalho et al., 2008 for an account of the laborious trial-and-error learning in tool-using chimpanzees). If all brains are surprisal-reducing machines, what is it about human brains that allows them to be guided so effectively, often foregoing laborious trial-and-error tweaking?

A common solution is to posit that humans evolved a special neural mechanism for re-representing information in a way which allows complex inferences, cognitive flexibility, language, (and self-awareness itself; Penn et al., 2008). The solution Clark offers – cursorily in the target article (§3.4, note xxxii) and more in depth in earlier work (e.g., Clark, 1998) – is that language together with other aspects of symbolic culture augment an otherwise un-remarkable pattern-completion, surprisal-reducing brain with faculties we have come to uniquely associate with the human mind, e.g.:

> "…linguistic formulation makes complex thoughts available to processes of mental attention. [It] enables us, for example, to pick out different elements of complex thoughts and to scrutinize each in turn" (Clark, 1998, p. 177–198).

Clark writes that "symbol-mediated loops" can "enable new forms of reentrant processing" (§3.4), but how does this work? Putting aside the question of how symbolic language and culture evolve in the first place, how might an agent's experience with symbols augment the prediction machinery? Answers to this question have tended to focus on (1) agent-level uses of language: explicit linguistic strategies such as verbal rehearsal, and mnemonic and chunking strategies (e.g., remembering an arbitrary sequence of letters by thinking of a sentence containing words that begin with those letters, or learning to tie a knot by thinking of a rabbit going in and out of a hole), and (2) explicit verbal mediation, i.e., "thinking in words." Indeed, this introspection of thinking in words is often so strong that it leads researchers to conflate that feeling of talking to oneself with the format of conceptual representations (Ryle, 1968; e.g., Carruthers, 2002; Levinson, 1997 for discussion).

This confusion can be clarified by considering the role language can play in generating top-down predictions (Lupyan, 2012a,b for discussion). A growing body of work suggests that language interfaces directly with the surprisal-reducing machinery at the core of predictive-coding models. Consider a task in which one hears an auditory cue (e.g., a barking sound) and then sees a picture (e.g., a dog). The goal is to respond "yes" if the cue and picture match at a conceptual level, and "no" otherwise (e.g., a car following a barking sound). The better the match between the top-down predictive signal and the bottom-up activation produced by the probe, the faster (or more accurately) subjects can respond. Lupyan and Thompson-Schill (2012) found that linguistic cues ("dog") were more effective than non-linguistic cues (e.g., a barking sound, a car horn), even though both cue types were judged as equally predictive and unambiguous of the associated category. As the delay between the cue and probe was increased, the difference between the verbal and non-verbal-cue conditions also increased. Under the influence of the label (through hypothesized top-down effects), the resultant representations appeared to become more similar across subjects with increasing delays in a way that they did not on trials without the verbal label. This provides a basic demonstration of how verbal

---

[1]It is tempting to assume that the inability to tell other animals what to do is a problem of communication. But the problem here is not communication, but that of flexibly activating task-relevant mappings. Language is not necessary for communicating instructions for a task as simple as this, but experience with language *may* be necessary for the ability to flexibly deploy task-relevant stimulus-response mappings without environmental guidance.

labels act as "cues" (Elman, 2009) altering how knowledge (e.g., of what a dog looks like) is brought online.

This effect of labels as "cues" – augmenting the processing of incoming sensory information – can also be observed in simple visual discrimination and even simple detection tasks (e.g., Lupyan, 2008; Lupyan and Spivey, 2010). Words appear to serve as especially efficient *category* cues to the system, selectively activating the features most typical/diagnostic of the target category, resulting in representations that allow more efficient discrimination between the target and non-target stimuli or between signal and noise (Ward and Lupyan, 2011). Indeed it is this that may be responsible for the facilitatory role labels appear to play in the learning of some novel categories (Lupyan et al., 2007; see Lupyan, 2012a for a computational model).

This approach of up- or down-regulating language can be used to partially overcome the limitation of not having access to human brains unaided by language[2]. Even small linguistic tweaks can augment ongoing processing even in apparently low-level perceptual tasks. By considering the functions language has on the predictive mech-

anisms of the brain, we can gain further insights not just in domains where language acts as a tool – allowing us to do such things as guide behavior by writing down cooking recipes – but on such a fundamental question as how it is that humans can tell each other what to do!

REFERENCES

Carruthers, P. (2002). The cognitive functions of language. *Behav. Brain Sci.* 25, 657–674.

Carvalho, S., Cunha, E., Sousa, C., and Matsuzawa, T. (2008). Chaînes opératoires and resource-exploitation strategies in chimpanzee (Pan troglodytes) nut cracking. *J. Hum. Evol.* 55, 148–163.

Clark, A. (1998). "Magic words: how language augments human computation," in *Language and Thought: Interdisciplinary Themes*, eds P. Carruthers and J. Boucher (New York, NY: Cambridge University Press), 162–183.

Clark, A. (in press). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.*

Elman, J. L. (2009). On the meaning of words and dinosaur bones: lexical knowledge without a lexicon. *Cogn. Sci.* 33, 547–582.

Levinson, S. C. (1997). "From outer to inner space: linguistic categories and non-linguistic thinking," in *Language and Conceptualization*, eds J. Nuyts and E. Pederson (Cambridge: Cambridge University Press), 13–45.

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

Lupyan, G. (2012a). "What do words do? Towards a theory of language-augmented thought," in *The Psychology of Learning and Motivation*, Vol. 57, ed. B. H. Ross (San Diego: Academic Press), 255–297. Available at: http://www.sciencedirect.com/science/article/pii/B9780123942937000078

Lupyan, G. (2012b). Linguistically modulated perception and cognition: the label-feedback hypothesis. *Front. Psychology* 3:54. doi: 10.3389/fpsyg.2012.00054

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

Lupyan, G., and Spivey, M. J. (2010). Making the invisible visible: auditory cues facilitate visual object detection. *PLoS ONE* 5, e11452. doi: 10.1371/journal.pone.0011452

Lupyan, G., and Thompson-Schill, S. L. (2012). The evocative power of words: activation of concepts by verbal and nonverbal means. *J. Exp. Psychol. Gen.* 141, 170–186.

Penn, D. C., Holyoak, K. J., and Povinelli, D. J. (2008). Darwin's mistake: explaining the discontinuity between human and nonhuman minds. *Behav. Brain Sci.* 31, 109–130.

Ryle, G. (1968). "A puzzling element in the notion of thinking," in *Studies in the Philosophy of Thought and Action*, ed. P. F. Strawson (Oxford: Oxford University Press), 7–23.

Spelke, E. S. (2003). "What makes us smart? Core knowledge and natural language," in *Language in Mind: Advances in the Study of Language and Thought*, eds D. Gentner and S. Goldin-Meadow (Cambridge, MA: MIT Press), 277–311.

Ward, E. J., and Lupyan, G. (2011). Linguistic penetration of suppressed visual representations. *J. Vis.* 11, 322.

---

[2]Special cases such as deaf children not exposed to Sign can be informative, but what is typically ignored is that such individuals nevertheless interact with linguistic beings and live in a world shaped by language and symbolic culture.