Authors: Arnon Lotem, Oren Kolodny, Joseph Y. Halpern, Luca Onnis, and Shimon Edelman

Abstract words: 61
Text words: 1,106
References words: 843
Total words: 1,949

The Bottleneck May Be the Solution, Not the Problem

Arnon Lotem
Dept. of Zoology, Tel Aviv University
Ramat Aviv, Israel

Oren Kolodny
Dept. of Biology, Stanford University
Stanford, CA, USA

Joseph Y. Halpern
Dept. of Computer Science, Cornell University
Ithaca, NY, USA

Luca Onnis
Division of Linguistics and Multilingual Studies, Nanyang Technological University
Singapore

Shimon Edelman
Dept. of Psychology, Cornell University
Ithaca, NY, USA

Abstract.

As a highly consequential biological trait, a memory "bottleneck" cannot escape selection pressures. It must therefore co-evolve with other cognitive mechanisms rather than act as an independent constraint. Recent theory and an implemented model of language acquisition suggest that a limit on working memory may evolve to help learning. Furthermore, it need not hamper the use of language for communication.

The target paper by Christiansen and Chater (C&C) makes many useful and valid observations about language that we happily endorse. Indeed, several of C&C's major points appear in our own papers, including: (i) the inability of non-chunked, "analog" approaches to language to compete with "digital" combinatorics over chunks (Edelman, 2008b); (ii) the centrality of chunking to modeling incremental, memory-constrained language acquisition and generation (Goldstein et al., 2010; Kolodny et al., 2015b) and the possible evolutionary roots of these features of language (Lotem and Halpern, 2012; Kolodny et al., 2014, 2015a); (iii) the realization that language experience has the form of a graph (Solan et al., 2005; cf. Edelman, 2008a, p.274), corresponding to C&C's "forest tracks" analogy; (iv) a proposed set of general principles for language acquisition and processing (Goldstein et al., 2010), one of which is essentially identical to C&C's "Now-or-Never Bottleneck." However, our theory is critically different in its causality structure. Rather than assuming that the memory limit is a fixed constraint to which all other traits must adapt, we view it as an adaptation that evolved to cope with computational challenges. Doing this brings it in line with standard practice in evolutionary biology, is more consistent with research findings, and raises numerous important research issues. We expand on these points below.

No biological trait can be simply assumed as a "constraint"

Viewing the "Now-or-Never Bottleneck" as an evolutionary constraint to which language adapts — C&C's central idea — is unwarranted. In evolutionary theory, biological constraints — as opposed to constraints imposed by physics and chemistry, which are not subject to biological evolution — cannot simply be assumed: they must be understood in terms of trade-offs among selective pressures. Clearly, birds wings evolved under aerodynamic constraints rather than vise versa. However, biological traits such as memory are not exempt from evolving. In proposing a bottleneck to which everything else in the system must adapt while the bottleneck itself remains fixed and independent (Figure 1 in the target article), C&C implicitly assume that it cannot evolve.

To justify this assumption, C&C should have offered evidence of stabilizing selection pressures that act against genetic variants coding for a broader or narrower bottleneck, and thereby affecting cognition and ultimately fitness. Alternatively, they might have assumed that the biological mechanisms underlying the memory bottleneck cannot be genetically variable — an odd assumption, which runs counter to substantial evidence in humans of (i) a range of verbal memory decay rates (Mueller and Krawitz, 2009), including in particular the longer verbal working memory span in individuals with Asperger's (Cui et al., 2010); (ii) heritable variation in language and in word memory (Stromswold, 2001; van Soelen et al., 2011) and in working memory (Blokland et al., 2011; Vogler et al., 2014); and (iii) variation in perceptual memory across species (Mery et al., 2007; Lind et al., 2015). Given that heritable variation in a trait means that it can respond to selection (e.g. Falconer, 1981), it is likely that the bottleneck *can* evolve, and that it is what it is because individuals with longer or shorter verbal working memory had lower biological fitness.[1]

If language is supported by domain-general mechanisms, verbal memory is even less immune to evolution

If the emergence of language constitutes a recent and radical departure from other cognitive phenomena, it is in principle possible that working memory evolved and stabilized prior to and separately from the "increasingly abstract levels of linguistic representation" posited by C&C. However, there are good arguments in support of a domain-general view of language (e.g., Chater and Christiansen, 2010). In particular, linguistic representations and processes are hardly as modular as C&C assume (Onnis and Spivey, 2012). Furthermore, theories of neural reuse (Anderson, 2010) point to the massive redeployment of existing mechanisms for new functions, resulting in brain regions coming to be involved in diverse cognitive functions. If circuits that support language continue contributing to non-linguistic functions (including working memory),

[1] If verbal memory indeed evolves, language is the niche in which it does so. The target paper seems to gloss over the intimate connection between cultural evolution and niche construction Odling-Smee et al., 2003. In focusing on how "linguistic patterns, which can be processed through that bottleneck, will be strongly selected," C&C ignore the possibility of there being also selection for individuals who can better process linguistic patterns.

a memory bottleneck is not a prior and independent constraint on language, but rather a trait that continues to evolve under multiple selective pressures, which include language.

The bottleneck may be the solution, not the problem

As we have suggested (Lotem and Halpern, 2008; Onnis et al., 2008; Goldstein et al., 2010; Lotem and Halpern, 2012), a limited working memory may be an adaptation for coping with the computational challenges involved in segmentation and network construction. (Importantly, regardless of whether this specific hypothesis is correct, entertaining such hypotheses is the only way of distinguishing a function from a constraint; cf. Stephens and Krebs, 1986, chapter 10.) A recently implemented model that includes this hypothesis has been tested on tasks involving language, birdsong, and foraging (Kolodny et al., 2014, 2015a; Menyhart et al., 2015; Kolodny et al., 2015b) The model includes a time window during which natural and meaningful patterns are likely to recur and thus to pass a test for statistical significance, while spurious patterns decay and are forgotten. We stress that rather than acting as a constraint, the duration of the window must co-evolve with the mechanisms influencing the distribution of data, so as to increase the effectiveness of memory representations (Lotem and Halpern, 2012).

We do agree with C&C regarding some of the consequences of the memory bottleneck, such as the need for online incremental construction of hierarchical representation. Indeed, our model effectively implements what C&C call "Chunk-and-Pass" (Kolodny et al., 2015b).[2] We believe, however, that the ultimate constraint on learning structure (such as that of language) in time and space is not the memory bottleneck in itself, but rather the computational challenges of chunking the data and of building hierarchies.

Biological communication is about affecting behavior, not pumping bits

Our final point focuses on the communicative function of language. Viewing a memory window as a communication "bottleneck" suggests that massive amounts of information must flow through the channel in question. However, the real function of a message is to influence the rich network of connotations and interconnections already present in the listener's brain (cf. Edelman, 2015, sec. 2.3). Communication is about generating adaptive behavioral changes (Burghardt, 1970; Green and Marler, 1979) — the listener gleans from it cues relevant to decision making. For this, a signal must be informative and reliable in the given context (Leger, 1993); the amount of information is not the main issue (except as a signal of quality, as in complex courtship songs; Lachmann et al., 2001). This implies that evolutionary selection in language is for how messages fit into the information already represented by their recipient; a bottleneck may not impose significant constraints here.

References

Anderson, M. L. (2010). Neural re-use as a fundamental organizational principle of the brain. *Behavioral and Brain Sciences*, 34:245–266.

[2]As C&C note, correctly, regarding Chunk-and-Pass, "it is entirely possible that linguistic input can simultaneously, and perhaps redundantly, be chunked in more than one way." This suggests that chunking on its own, especially when carried out recursively/hierarchically, is likely to severely exacerbate the combinatorial problem faced by the learner, rather than resolve the bottleneck issue.

Blokland, G. A. M., McMahon, K. L., Thompson, P. M., Martin, N. G., de Zubicaray, G. I., and Wright, M. J. (2011). Heritability of working memory brain activation. *The Journal of Neuroscience*, 31:10882–10890.

Burghardt, G. M. (1970). Defining 'communication'. In Johnston Jr., J. W., Moulton, D. G., and Turk, A., editors, *Communication by Chemical Signals*. Century-Crofts, New York.

Chater, N. and Christiansen, M. H. (2010). Language acquisition meets language evolution. *Cognitive Science*, 34:1131–1157.

Cui, J., Gao, D., Chen, Y., Zou, X., and Wang, Y. (2010). Working memory in early-school-age children with Asperger's syndrome. *J. Autism Dev. Disord.*, 40:958–967.

Edelman, S. (2008a). *Computing the mind: how the mind really works*. Oxford University Press, New York, NY.

Edelman, S. (2008b). On the nature of minds, or: Truth and consequences. *Journal of Experimental and Theoretical AI*, 20:181–196.

Edelman, S. (2015). The minority report: some common assumptions to reconsider in the modeling of the brain and behavior. *Journal of Experimental and Theoretical Artificial Intelligence*, 27:–. in press.

Falconer, D. S. (1981). *Introduction to quantitative genetics*. Longman.

Goldstein, M. H., Waterfall, H. R., Lotem, A., Halpern, J., Schwade, J., Onnis, L., and Edelman, S. (2010). General cognitive principles for learning structure in time and space. *Trends in Cognitive Sciences*, 14:249–258.

Green, S. and Marler, P. (1979). The analysis of animal communication. In Marler, P. and Vandenbergh, J. G., editors, *Handbook of behavioral neurobidlogy: Vol.3. Social behavior and communication*, pages 73–158. Plenum Press, New York.

Kolodny, O., Edelman, S., and Lotem, A. (2014). The evolution of continuous learning of the structure of the environment. *Journal of the Royal Society Interface*, 11:20131091.

Kolodny, O., Edelman, S., and Lotem, A. (2015a). Evolution of protolinguistic abilities as a byproduct of learning to forage in structured environments. *Proceedings of the Royal Society of London B*. In press.

Kolodny, O., Lotem, A., and Edelman, S. (2015b). Learning a generative probabilistic grammar of experience: a process-level model of language acquisition. *Cognitive Science*, 39:227–267.

Lachmann, M., Számadó, S., and Bergstrom, C. T. (2001). Cost and conflict in animal signals and human language. *Proceedings of the National Academy of Science*, 98:13189–13194.

Leger, D. W. (1993). Contextual sources of information and responses to animal communication signals. *Psychological Bulletin*, 113:295–304.

Lind, J., Enquist, M., and Ghirlanda, S. (2015). Animal memory: A review of delayed matching-to-sample data. *Behavioural Processes*. In press.

Lotem, A. and Halpern, J. Y. (2012). Coevolution of learning and data-acquisition mechanisms: a model for cognitive evolution. *Phil. Trans. R. Soc. B*, 367:2686–2694.

Lotem, A. and Halpern, Y. J. (2008). A data-acquisition model for learning and cognitive development and its implications for autism. Computing and information science technical reports, Cornell University. Available online at http://hdl.handle.net/1813/10178.

Menyhart, O., Kolodny, O., Goldstein, M. H., Devoogd, T., and Edelman, S. (2015). Juvenile zebra finches learn the underlying statistical regularities in their father's song. *Frontiers in Psychology*, 6:571.

Mery, F., Belay, A. T., So, A. K., Sokolowski, M. B., and Kawecki, T. J. (2007). Natural polymorphism affecting learning and memory in Drosophila. *Proceedings of the National Academy of Science*, 104:13051–13055.

Mueller, S. T. and Krawitz, A. (2009). Reconsidering the two-second decay hypothesis in verbal working memory. *Journal of Mathematical Psychology*, 53:14–25.

Odling-Smee, F. J., Laland, K. N., and Feldman, M. W. (2003). *Niche construction: the neglected process in evolution*, volume MPB 37. Princeton University Press, Princeton, NJ.

Onnis, L. and Spivey, M. J. (2012). Toward a new scientific visualization for the language sciences. *Information*, 3:124–150.

Onnis, L., Waterfall, H. R., and Edelman, S. (2008). Learn locally, act globally: Learning language from variation set cues. *Cognition*, 109:423–430.

Solan, Z., Horn, D., Ruppin, E., and Edelman, S. (2005). Unsupervised learning of natural languages. *Proceedings of the National Academy of Science*, 102:11629–11634.

Stephens, D. W. and Krebs, J. R. (1986). *Foraging Theory*. Princeton University Press, Princeton, NJ.

Stromswold, K. (2001). The heritability of language: A review and metaanalysis of twin, adoption, and linkage studies. *Language*, 77:647–723.

van Soelen, I. L. C., Brouwer, R. M., van Leeuwen, M., Kahn, R. S., Hulshoff Pol, H. E., and Boomsma, D. I. (2011). Heritability of verbal and performance intelligence in a pediatric longitudinal sample. *Twin Research and Human Genetics*, 14:119–128.

Vogler, C., Gschwind, L., Coyne, D., Freytag, V., Milnik, A., Egli, T., Heck, A., de Quervain, D. J., and Papassotiropoulos, A. (2014). Substantial SNP-based heritability estimates for working memory performance. *Translational Psychiatry*, 4:e438.