City Research Online

City, University of London Institutional Repository

This is the accepted version of the paper.

This version of the publication may differ from the final published version.

Linguistics, cognitive psychology, and the now-or-never bottleneck

Ansgar Endress
Department of Psychology, City University, London UK

Roni Katzir
Department of Linguistics and Sagol School of Neuroscience, Tel Aviv University, Israel

Ansgar Endress
Department of Psychology
City University, London
UK
Email: ansgar.endress.1@city.ac.uk
Phone: +44 (0)20 7040 8640

Abstract

Christiansen & Chater (CC)'s key premise is that "if linguistic information is not processed rapidly, that information is lost for good". From this "Now-or-Never Bottleneck" (NNB), CC derive "wide-reaching and fundamental implications for language processing, acquisition and change as well as for the structure of language itself". We question both the premise and the consequentiality of its purported implications.

Problematic premises
CC base the NNB on the observation that sensory memory disappears quickly in explicit memory tasks. We note, first, that not all forms of explicit memory are short-lived. For example, children remember words encountered once after a month (Carey & Bartlett, 1978; Markson & Bloom, 1997). More importantly, it is by no means clear that explicit memory is the (only) relevant form of memory for language processing and acquisition, nor how quickly other forms of memory decay. For example, the perceptual learning literature suggests that learning can occur even in the absence of awareness of the stimuli (Watanabe et al., 2001; Seitz & Watanabe, 2003), and sometimes has long-lasting effects (Schwab et al., 1985). Similarly, visual memories that start decreasing over a few seconds can be stabilized by presenting items another time (Endress & Potter, 2014). At a minimum, then, such memory traces are long-lasting enough for repeated exposure to have cumulative learning effects.

Information that is not even perceived is thus used for learning and processing, and some forms of memory do not disappear immediately. Hence, it is still an open empirical question whether poor performance in explicit recall tasks provides severe constraints on processing and learning.

We note, in passing, that even if relevant forms of memory were short-lived, this would not necessarily be a bottleneck. Mechanisms to make representations last longer such as self-sustained activity are well documented in many brain regions (Major & Tank, 2004), and one might assume that memories can be longer-lived when this is adaptive. Short-lived memories might thus be an adaptation rather than a bottleneck (e.g., serving to reduce information load for various computations).

Problematic 'implications'
CC use the NNB to advance the following view: language is a skill (specifically, the skill of parsing predictively); this skill is what children acquire (rather than some theory-like knowledge); and there are few if any restrictions on linguistic diversity. CC's conclusions do not follow from the NNB and are highly problematic. Below, we discuss some of the problematic inferences regarding processing, learning and evolution.

Regarding processing, CC claim that the NNB implies that knowledge of language is the skill of parsing predictively. There is indeed ample evidence for a central role for prediction in parsing (e.g., Levy 2008), but this is not a consequence of the NNB: the advantages of predictive processing are orthogonal to the NNB, and, even assuming the NNB, processing might still occur element by element without predictions. CC also claim that the NNB implies a processor with no explicit representation of syntax (other than what can be read off the parsing process as a trace). It is unclear what they actually mean with this claim, though. First, if CC mean that the parser does not construct full syntactic trees but rather a minimum that allows semantics and phonology to operate, they just echo a view discussed by Pulman (1986) and others. While this view is an open possibility, we do not see how it follows from the NNB. Second, if CC mean that the NNB implies that parsing does not use explicit syntactic knowledge, this view is incorrect: many parsing algorithms (e.g., LR, Earley's algorithm, incremental CKY) respect the NNB by being incremental and not needing to refer back to raw data (they can all refer to the result of earlier processing instead) and yet make reference to explicit syntax. Finally, we note that prediction-based parser-only models in the literature that do not incorporate explicit representations of syntactic structure (e.g., Elman,

1990; McCauley & Christiansen, 2011) fail to explain why we can recognize unpredictable sentences as grammatical (e.g., *Evil unicorns devour xylophones*).

Regarding learning, CC claim that the NNB is incompatible with approaches to learning that involve elaborate linguistic knowledge. This, however, is incorrect: the only implication of the NNB for learning is that if memory is indeed fleeting, any learning mechanism must be online rather than batch, relying only on current information. But online learning does not rule out theory-based models of language in any way (e.g., Börschinger & Johnson 2011). In fact, it has been argued that online variants of theory-based models provide particularly good approximations to empirically observed patterns of learning (e.g., Frank et al. 2010).

Regarding the evolution of language (which CC conflate with the biological evolution of language), they claim that it is item-based and gradual, and that linguistic diversity is the norm, with few if any true universals. However, it is unclear how these claims might follow from the NNB, and they are inconsistent with the relevant literature. For example, language change has been argued to be abrupt and nonlinear (see Niyogi & Berwick 2009), often involving what look like changes in abstract principles rather than concrete lexical items. As for linguistic diversity, CC repeat claims in Christiansen & Chater 2008 and Evans & Levinson 2009, but those works ignore the strongest typological patterns revealed by generative linguistics. For example, no known language allows for a single conjunct to be displaced in a question (Ross 1967): we might know that Kim ate *peas and something* yesterday and wonder what that *something* is, but in no language can we use a question of the form *What did Kim eat peas and yesterday? to* inquire about it. Likewise, in *Why did John wonder who Bill hit?*, one can only ask about the cause of the wondering, not of the hitting (see Huang 1982, Rizzi 1990). Typological data thus reveal significant restrictions on linguistic diversity.

Conclusion
Language is complex. Our efforts to comprehend it are served better by detailed analysis of the cognitive mechanisms at our disposal than by grand theoretical proposals that ignore the relevant psychological, linguistic and computational distinctions.

References
Börschinger, B. and Johnson, M. (2011). A particle filter algorithm for Bayesian word segmentation. In Proceedings of the Australasian Language Technology Association Workshop, pages 10–18.

Carey, S. & Bartlett, E. (1978). Acquiring a single word. *Papers and Reports on Child Language Development*, 15, 17-29

Christiansen, M. H. and Chater, N. (2008). Language as shaped by the brain. Behavioral and Brain Sciences, 31(05):489–509.

Elman, J. L. (1990). Finding structure in time. Cognitive Science, 14(2):179–211.

Endress, A. D. & Potter, M. C. (2014). Something from (almost) nothing: Buildup of object memory from forgettable single fixations. *Atten Percept Psychophys, 76*, 2413-2423

Evans, N. and Levinson, S. (2009). The myth of language universals: Language diversity and its importance for cognitive science. Behavioral and Brain Sciences, 32:429–492.

Frank, M., Goldwater, S., Griffiths, T., and Tenenbaum, J. (2010). Modeling human performance in statistical word segmentation. Cognition, 117:107–125.

Huang, C.-T. J. (1982). Logical relations in Chinese and the theory of grammar. PhD thesis, Massachusetts Institute of Technology.

Levy, R. (2008). Expectation-based syntactic comprehension. Cognition, 106:1126–1177.

Major, G. and Tank, D. (2004). Persistent neural activity: prevalence and mechanisms. *Current Opinion in Neurobiology*, 14(6):675–684.

Markson, L. & Bloom, P. (1997). Evidence against a dedicated system for word learning in children. *Nature*, 385, 813 - 815

McCauley, S. M. and Christiansen, M. H. (2011). Learning simple statistics for language comprehension and production: The cappuccino model. In Proceedings of the 33rd Annual Conference of the Cognitive Science Society, pages 1619–1624.

Niyogi, P. and Berwick, R. C. (2009). The proper treatment of language acquisition and change in a population setting. Proceedings of the National Academy of Sciences, 106:10124– 10129.

Pulman, S. G. (1986). Grammars, parsers, and memory limitations. Language and Cognitive Processes, 1(3):197–225.

Rizzi, L. (1990). Relativized Minimality. MIT Press, Cambridge, Massachusetts.

Ross, J. R. (1967). Constraints on Variables in Syntax. PhD thesis, MIT, Cambridge, MA.

Schwab, E. C.; Nusbaum, H. C. & Pisoni, D. B. (1985). Some effects of training on the perception of synthetic speech. *Human Factors, 27*, 395-408

Seitz, A. R. & Watanabe, T. (2003). Psychophysics: Is subliminal learning really passive? *Nature, 422*, 36

Watanabe, T., Náñez, J. E. & Sasaki, Y. (2001). Perceptual learning without perception. *Nature, 413*, 844-848