

Commentary on:

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

Citation of this commentary:

Farmer, T. A., Brown, M., & Tanenhaus, M. K. (2013). Prediction, explanation, and the role of generative models in language processing. *Behavioral and Brain Sciences*, 36, 211-212.

Prediction, explanation, and the role of generative models in language processing

Thomas A. Farmer^{a,b}, Meredith Brown^a and Michael K. Tanenhaus^a

^a*Department of Brain and Cognitive Sciences and* ^b*Center for Language Sciences, University of Rochester, Rochester, NY 14627-0268.*

tfarmer@bcs.rochester.edu

mbrown@bcs.rochester.edu

mtan@bcs.rochester.edu

Abstract: We propose, following Clark, that generative models also play a central role in the perception and interpretation of linguistic signals. The data explanation approach provides a rationale for the role of prediction in language processing and unifies a number of phenomena, including multiple-cue integration, adaptation effects, and cortical responses to violations of linguistic expectations.

Traditional models of language comprehension assume that language processing involves recognizing patterns, for example, words, by mapping the signal onto existing

representations, retrieving information associated with these stored representations, and then using rules based on abstract categories (e.g., syntactic rules) to build structured representations. Four aspects of the literature are inconsistent with this framework. First, listeners are exquisitely sensitive to fine-grained, sub-categorical properties of the signal, making use of this information rather than discarding it (McMurray et al. 2009). Second, comprehenders rapidly integrate constraints at multiple grains. Third, they generate expectations about likely input at multiple levels of representation. Finally, adaptation is ubiquitous in language processing. These results can be unified if we assume that comprehenders use internally generated predictions at multiple levels to *explain* the source of the input, and that prediction error is used to update the generative models in order to facilitate more accurate predictions in the future.

Extended to the domain of language processing, Clark's framework predicts that expectations at higher levels of representation (e.g., syntactic expectations) should constrain interpretation at lower levels of representation (e.g., speech perception). According to this view, listeners develop fine-grained probabilistic expectations about how lexical alternatives are likely to be realized in context (e.g., *net* vs. *neck*) that propagate from top to bottom through the levels of a hierarchically organized system representing progressively more fine-grained perceptual information. Provisional hypotheses compete to explain the data at each level, with the predicted acoustic realization of each alternative being evaluated against the actual form of the input, resulting in a residual feed-forward error signal propagated up the hierarchy. As the signal unfolds, then, the activation of a particular lexical candidate should be inversely proportional to the joint error signal at all levels of the hierarchy (i.e., the degree of

divergence between the predicted acoustic realization of that candidate and the actual incoming signal), such that candidate words whose predicted realizations are most congruent with the acoustic signal are favored.

Hierarchical predictive processing therefore provides a potential explanatory framework for understanding a wide variety of context effects and cue integration phenomena in spoken word recognition. Converging evidence suggests that the initial moments of competition between lexical alternatives are constrained by multiple sources of information from different dimensions of the linguistic input (e.g., Dahan & Tanenhaus 2004; Kukona et al. 2011), including information external to the linguistic system, such as visually conveyed social information (Hay & Drager 2010; Staum Casasanto 2008) and high-level information about a speaker's linguistic ability (Arnold et al. 2007). Crucially, lexical processing is influenced by information preceding the target word by several syllables or clauses (Dilley & McAuley 2008; Dilley & Pitt 2010) and this information affects listeners' expectations (Brown et al. 2011; 2012). The integration of these various constraints, despite their diversity, is consistent with the hypothesis that disparate sources of constraint are integrated within generative models in the language processing system.

Clark's framework also helps explain a recent set of results on context effects in reading that are surprising from the viewpoint of more traditional theories that emphasize the bottom-up, feed-forward flow of information. Farmer et al. (2006) demonstrated that when a sentential context conferred a strong expectation for a word of a given

grammatical category (as in *The child saved the...*, where a noun is strongly expected), participants were slower to read the incoming noun when the *form* of it (i.e., its phonological/orthographic properties) was atypical with respect to other words in the expected category. In a subsequent MEG experiment, Dikker et al. (2010) showed that at about 100 msec post-stimulus onset – timing that is unambiguously associated with perceptual processing – a strong neural response was elicited when there was a mismatch between form and syntactic expectation. Moreover, the source of the effect was localized to the occipital lobe, suggesting that the visual system had access to syntactic representations. These results provide support for Clark’s hypothesis that “if the predictive processing story is correct, we expect to see powerful context effects propagating quite low down the processing hierarchy” (sect. 3.1, para. 8). Linguistic context is used to generate expectations about form-based properties of upcoming words, and these expectations are propagated to perceptual cortices (Tanenhaus & Hare 2007).

This framework also serves to specify the functionality of the prediction error that arises when some degree of mismatch between a prediction and the incoming signal occurs. In behavioral and Event-Related Potential (ERP) experiments, prediction-input mismatch frequently results in increased processing difficulty, typically interpreted as evidence that prediction is being made. But, under Clark’s framework, the error signal assumes functionality; in part, it serves to adjust higher-level models such that they better approximate future input. The explanatory power of this hypothesis can best be seen when considering the large amount of relatively recent literature on adaptation within linguistic domains. Whether in the domain of speech perception (Kleinschmidt & Jaeger

2011; Kraljic et al. 2008), syntactic processing (Farmer et al. 2011; Fine et al. under review; Wells et al. 2009), prosody (Kurumada et al. 2012), or pragmatics (Grodner & Sedivy 2011), it has become increasingly apparent that readers and listeners continually update their expectations about the likelihood of encountering some stimulus based on their exposure to the statistical regularities of a specific experimental context. Adaptation of expectations is predicted by Clark's framework, and it may be taken as evidence that prediction-input mismatch produces an error signal that is fed forward to update the relevant generative models.

In sum, Clark's hierarchical prediction machine hypothesis provides a framework that we believe will unify the literature on prediction in language processing. This unification will necessarily involve systematic examination of what aspects of the stimulus are predicted, when in the chain of processing these predictions are generated and assessed, and the precise form of these generative models. This task will be challenging because it is likely that generative models use signal-relevant properties that do not map to the standard levels of linguistic representation that are incorporated into most models of language processing.

References

- Arnold, J. E., Hudson, C. L. & Tanenhaus, M. K. (2007) If you say thee uh you are describing something hard: The on-line attribution of disfluency during reference comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 33:914–30.
- Brown, M., Dilley, L. C. & Tanenhaus, M. K. (2012) Real-time expectations based on

- context speech rate can cause words to appear or disappear. In: *Proceedings of the 34th Annual Conference of the Cognitive Science Society*, ed. N. Miyake, D. Peebles, & R. P. Cooper, pp. 1374–79. Cognitive Science Society.
- Brown, M., Salverda, A. P., Dilley, L. C. & Tanenhaus, M. K. (2011) Expectations from preceding prosody influence segmentation in online sentence processing. *Psychonomic Bulletin and Review* 18:1189–96.
- Dahan, D. & Tanenhaus, M. K. (2004) Continuous mapping from sound to meaning in spoken-language comprehension: Evidence from immediate effects of verb-based constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 30:498–513.
- Dikker, S., Rabagliati, H., Farmer, T. A. & Pytkkanen, L. (2010) Early occipital sensitivity to syntactic category is based on form typicality. *Psychological Science* 21:629–34.
- Dilley, L. C. & McAuley, J. D. (2008) Distal prosodic context affects word segmentation and lexical processing. *Journal of Memory and Language* 59:294–311.
- Dilley, L. C. & Pitt, M. (2010) Altering context speech rate can cause words to appear or disappear. *Psychological Science* 21:1664–70.
- Farmer, T. A., Christiansen, M. H. & Monaghan, P. (2006) Phonological typicality influences on-line sentence comprehension. *Proceedings of the National Academy of Sciences USA* 103:12203–208.
- Farmer, T. A., Monaghan, P., Misyak, J. B. & Christiansen, M. H. (2011) Phonological typicality influences sentence processing in predictive contexts: A reply to Staub et al. (2009). *Journal of Experimental Psychology: Learning, Memory, and*

Cognition 37:1318–25.

- Fine, A. B., Jaeger, T. F., Farmer, T. A. & Qian, T. (under review). Rapid expectation adaptation during syntactic comprehension.
- Grodner, D. & Sedivy, J. (2011) The effect of speaker-specific information on pragmatic inferences. In: *The processing and acquisition of reference*, vol. 2327, ed. E. Gibson & N. Pearlmuter, pp. 239–72. MIT Press.
- Hay, J. & Drager, K. (2010) Stuffed toys and speech perception. *Linguistics* 48:865–92.
- Kleinschmidt, D. & Jaeger, T. F. (2011) *A Bayesian belief updating model of phonetic recalibration and selective adaptation*. Association for Computational Linguistics—Computational Modeling and Computational Linguistics.
- Kraljic, T., Samuel, A. G. & Brennan, S. E. (2008) First impressions and last resorts: How listeners adjust to speaker variability. *Psychological Science* 19:332–38.
- Kukona, A., Fang, S., Aicher, K. A., Chen, H. & Magnuson, J. S. (2011) The time course of anticipatory constraint integration. *Cognition* 119:23–42.
- Kurumada, C., Brown, M. & Tanenhaus, M. K. (2012) Pragmatic interpretation of contrastive prosody: It looks like adaptation. In: *Proceedings of the 34th Annual Conference of the Cognitive Science Society*, ed. N. Miyake, D. Peebles, & R. P. Cooper, pp. 647–52. Cognitive Science Society.
- McMurray, B., Tanenhaus, M. K. & Aslin, R. N. (2009) Within-category VOT affects recovery from “lexical” garden paths: Evidence against phoneme-level inhibition. *Journal of Memory and Language* 60:65–91.
- Stam Casasanto, L. (2008) Does social information influence sentence processing? In: *Proceedings of the 30th Annual Conference of the Cognitive Science Society*, ed.

B. C. Love, K. McRae, & V. M. Sloutsky, pp. 799–804. Cognitive Science Society.

Tanenhaus, M. K. & Hare, M. (2007) Phonological typicality and sentence processing. *Trends in Cognitive Science* 11:93–95.

Wells, J. B., Christiansen, M. H., Race, D. S., Acheson, D. J. & MacDonald, M. C. (2009) Experience and sentence comprehension: Statistical learning and relative clause comprehension. *Cognitive Psychology* 58:250–71.