Published in final edited form as:

Behav Brain Sci. 2013 August ; 36(4): 359–360. doi:10.1017/S0140525X12002762.

Seeking predictions from a predictive framework

T. Florian Jaeger^{a,b} and Victor Ferreira^c

^aDepartment of Brain and Cognitive Sciences, University of Rochester, Rochester, NY 14627-0268

^bDepartment of Computer Science, University of Rochester, Rochester, NY 14627

^cDepartment of Psychology 0109, University of California, San Diego, La Jolla, CA 92093-0109

Abstract

We welcome the proposal to use forward models to understand predictive processes in language processing. However, Pickering & Garrod (P&G) miss the opportunity to provide a strong framework for future work. Forward models need to be pursued in the context of learning. This naturally leads to questions about *what* prediction error these models aim to minimize.

Pickering & Garrod (P&G) are not the first to propose that comprehension is a predictive process (e.g., Hale 2001; Levy 2008; Ramsar et al. 2010). Similarly, recent work has found that language production is sensitive to prediction in ways closely resembling comprehension (e.g., Aylett & Turk 2004; Jaeger 2010). We believe that forward models (1) offer an elegant account of prediction effects and (2) provide a framework that could generate novel predictions and guide future work. However, in our view, the proposal by P&G fails to advance either goal because it does not take into account two important properties of forward models. The first is learning; the second is the nature of the prediction error that the forward model is minimizing.

Learning

Forward models have been a successful framework for motor control in large part because they provide a unifying framework, not only for prediction, but also for learning. Since their inception, forward models have been used to study learning – both acquisition and adaptation throughout life. However, except for a brief mention of “tuning” (target article, sect. 3.1, para. 15), P&G do not discuss what predictions their framework makes for implicit learning during language production, despite the fact that construing language processing as prediction in the context of learning readily explains otherwise puzzling findings from production (e.g., Roche et al. 2013; Warker & Dell 2006), comprehension (e.g., Clayards et al. 2008; Farmer et al. 2013; Kleinschmidt et al. 2012) and acquisition (Ramsar et al. 2010). If connected to learning, forward models can explain *how we learn to align our predictions* during dialogue (i.e., learning in order to reduce future prediction errors, Fine et

al., submitted; Jaeger & Snider 2013; for related ideas, see also Chang et al. 2006; Fine & Jaeger 2013; Kleinschmidt & Jaeger 2011; Sonderegger & Yu 2010).

Prediction errors

Deriving testable predictions from forward models is integrally tied to the nature of the prediction error that the system is meant to minimize during self- and other-monitoring (i.e., the function of the model, cf. Guenther et al. 1998). P&G do not explicitly address this. They do, however, propose separate forward models at all levels of linguistic representations. These forward models seem to have just one function, to predict the perceived *linguistic unit* at each level. For example, the syntactic forward model predicts the “syntactic percept,” which is used to decide whether the production plan needs to be adjusted (how this comparison proceeds and what determines its outcome is left unspecified).

Minimizing communication error: A proposal

If one of the goals of language production is to be understood – or even to communicate the intended message both robustly and efficiently (Jaeger 2010; Lindblom 1990) – correctly predicting the intended linguistic units should only be relevant *to the extent that not doing so impedes being understood*. Therefore, the prediction error that forward models in production should aim to minimize is not the perception of linguistic units, but the outcome of the entire inference process that constitutes comprehension. Support for this alternative view comes from work on motor control, work on articulation, and cross-linguistic properties of language.

For example, if the speaker produces an underspecified referential expression but is understood, there is no need to self-correct (as observed in research on conceptual pacts, Brennan & Clark 1996). This view would explain why only reductions of words with low confusability tend to enter the lexicon (e.g., “strodney,” rather than “extrary,” for “extraordinary”). If, however, the function of the forward model is to predict linguistic units, as P&G propose, no such generalization is expected. Rather, *any* deviation from the target phonology will cause a prediction error, regardless of whether it affects the likelihood of being understood. Similar reasoning applies to the reduction of morphosyntactic units, which often is blocked when it would cause systemic ambiguity (e.g., differential or optional case-marking, Fedzechkina et al. 2012; see also Ferreira 2008).

Research on motor control finds that not all prediction errors are created equal: Stronger adaptation effects are found after task-relevant errors (Wei & Körding 2009). Indeed, in a recent perturbation study on production, Frank (2011) found that speakers exhibit stronger error correction if the perceived deviation from the intended acoustics makes the actual production more similar to an existing word (see also Perkell et al. 2004).

This view also addresses another shortcoming of P&G’s proposal. At several points, P&G state that the forward models make impoverished predictions. Perhaps predictions are impoverished only in that they map the efference copy directly onto the predicted meaning (rather than the intermediate linguistic units).

Of course, the goal of reducing the prediction error for efficient information transfer is achieved by reducing the prediction error at the levels assumed by P&G. In this case, the architecture assumed by P&G would *follow* from the more general principle described here. However, in a truly predictive learning framework (Clark 2013), there is no guarantee that the levels of representation that such models would learn in order to minimize prediction errors would neatly map onto those traditionally assumed (cf. Baayen et al. 2011).

Finally, we note that, in the architecture proposed by P&G, the production forward model seems to serve no purpose but to be the input of the comprehension forward model (sect. 3.1, Fig. 5; sect. 3.2, Fig. 6). Presumably, the output of, for example, the syntactic production forward model will be a syntactic plan. Hence, the syntactic comprehension forward model takes syntactic plans as input. The *output* of that comprehension forward model must be akin to a parse, as it is compared to the output of the actual comprehension model. Neither of these components seems to fulfill any independent purpose. Why not map straight from the syntactic efference copy to the predicted “syntactic percept”? If forward models are used as a computational framework, rather than as metaphor, one of their strengths is that they can map efference copies *directly* onto the reference frame that is required for effective learning and minimization of the relevant prediction error (cf. Guenther et al. 1998).

References

- Aylett M, Turk A. The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*. 2004; 47(1):31–56. [PubMed: 15298329]
- Baayen RH, Milin P, Filipovi D, Durdevi D, Hendrix P, Marelli M. An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review*. 2011; 118:438–482. [PubMed: 21744979]
- Brennan SE, Clark HH. Conceptual pacts and lexical choice in conversation. *Learning, Memory*. 1996; 22(6):1482–1493.
- Chang F, Dell GS, Bock JK. Becoming Syntactic. *Psychological Review*. 2006; 113(2):234. [TFJ]. [PubMed: 16637761]
- Clark A. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Brain and Behavioral Sciences*. 2013; 36(3):181–253.
- Clayards M, Tanenhaus MK, Aslin RN, Jacobs RA. Perception of speech reflects optimal use of probabilistic speech cues. *Cognition*. 2008; 108(3):804–809. [PubMed: 18582855]
- Farmer T, Brown M, Tanenhaus MK. Prediction, explanation, and the role of generative models in language processing. *Brain and Behavioral Sciences*. 2013; 36(3):211–212.
- Fedzechkina M, Jaeger TF, Newport E. Language learners restructure their input to facilitate efficient communication. *Proceedings of the National Academy of Sciences of the United States of America*. 2012; 109(44):17897–17902. [PubMed: 23071337]
- Ferreira VS. Ambiguity, accessibility, and a division of labor for communicative success. *Psychology of Learning and Motivation*. 2008; 49:209–246.
- Fine AB, Jaeger TF. Evidence for implicit learning in syntactic comprehension. *Cognitive Science*. 2013; 37(3):578–591. [PubMed: 23363004]
- Fine AB, Jaeger TF, Farmer T, Qian T. Rapid linguistic adaptation during syntactic comprehension. (submitted).
- Frank, A. Integrating linguistic, motor, and perceptual information in language production. University of Rochester; 2011.

- Guenther FH, Hampson M, Johnson D. A theoretical investigation of reference frames for the planning of speech movements. *Psychological Review*. 1998; 105(4):611. [PubMed: 9830375]
- Hale, J. North American Chapter of the Association for Computational Linguistics (NAACL). Association for Computational Linguistics; 2001. A probabilistic early parser as a psycholinguistic model; p. 1-8.
- Jaeger TF. Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*. 2010; 61:23–62. [PubMed: 20434141]
- Jaeger TF, Snider N. Alignment as a consequence of expectation adaptation: Syntactic priming is affected by the prime's prediction error given both prior and recent experience. *Cognition*. 2013; 127(1):57–83. [PubMed: 23354056]
- Kleinschmidt, D.; Fine, AB.; Jaeger, TF. A belief-updating model of adaptation and cue combination in syntactic comprehension. *Proceedings of the 34rd Annual Meeting of the Cognitive Science Society (CogSci12)*; 2012. p. 605-610.
- Kleinschmidt, D.; Jaeger, TF. A Bayesian belief updating model of phonetic recalibration and selective adaptation. *Proceedings of the Cognitive Modeling and Computational Linguistics Workshop at ACL*; June 23rd; Portland, OR. 2011. p. 10-19.
- Levy R. Expectation-based syntactic comprehension. *Cognition*. 2008; 106(3):1126–1177. [PubMed: 17662975]
- Lindblom B. Explaining phonetic variation: A sketch of the H&H theory. *Speech production and speech modelling*. 1990; 55:40339.
- Perkell JS, Guenther FH, Lane H, Matthies ML, Stockmann E, Tiede M, Zandipour M. The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts. *The Journal of the Acoustical Society of America*. 2004; 116:2338. [PubMed: 15532664]
- Ramscar M, Yarlett D, Dye M, Denny K, Thorpe K. The effects of feature-label-order and their implications for symbolic learning. *Cognitive Science*. 2010; 34(6):909–957. [PubMed: 21564239]
- Roche, J.; Dale, R.; Kreuz, RJ.; Jaeger, TF. Learning to avoid syntactic ambiguity. Ms., University of Rochester; 2013.
- Sonderegger, M.; Yu, A. A rational account of perceptual compensation for coarticulation. Paper presented at the *Proceedings of the 32nd Annual Meeting of the Cognitive Science Society (CogSci10)*; 2010.
- Warker JA, Dell GS. Speech errors reflect newly learned phonotactic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2006; 32(2):387.
- Wei K, Körding K. Relevance of error: What drives motor adaptation? *Journal of neurophysiology*. 2009; 101(2):655–664. [PubMed: 19019979]