*Rochester, Rochester, NY 14627;* [c]*Department of Psychology 0109, University of California, San Diego, La Jolla, CA 92093-0109.*
**fjaeger@bcs.rochester.edu** **vferreira@ucsd.edu**
**http://www.hlp.rochester.edu/** **http://lpl.ucsd.edu/**

**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; Ramscar 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 (Ramscar 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., "strodny," 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).

## Prediction plays a key role in language development as well as processing

doi:10.1017/S0140525X12002609

Matt A. Johnson,[a] Nicholas B. Turk-Browne,[a] and Adele E. Goldberg[b]
[a]*Department of Psychology, Princeton University, Princeton, NJ 08544;* [b]*Program in Linguistics, Princeton University, Princeton, NJ 08544.*
**majthree@princeton.edu** **ntb@princeton.edu**
**adele@princeton.edu**
**www.princeton.edu/ntblab** **www.princeton.edu/~adele**

**Abstract:** Although the target article emphasizes the important role of prediction in language *use*, prediction may well also play a key role in the initial formation of linguistic representations, that is, in language *development*. We outline the role of prediction in three relevant language-learning domains: transitional probabilities, statistical preemption, and construction learning.

Pickering & Garrod (P&G) argue forcefully that language production and language comprehension are richly interwoven, allowing for fluid, highly interactive discourse to unfold. They note that a key feature of language that makes such fluidity possible is the pervasive use of prediction. Speakers predict and monitor their own language as they speak, allowing them to plan ahead and self-correct, and listeners predict upcoming utterances as they listen. The authors in fact provide evidence for predictive strategies at every level of language use: from phonology, to lexical semantics, syntax, and pragmatics.

Given the ubiquity of prediction in language *use*, an interesting consideration that P&G touch on only briefly is how prediction may be involved in the initial formation of linguistic representations, that is, in language *development*. Indeed, the authors draw heavily from forward modeling, invoking the Wolpert models as a possible schematic for their dynamic, prediction-based system. And although their inclusion is surely appropriate for discourse and language use, these models are fundamentally models of *learning* (e.g., Wolpert 1997; Wolpert et al. 2001). Hence, the degree to which our predictions are fulfilled (or violated) might have enormous consequences for linguistic representations and, ultimately, for the predictions we make in the future. More generally, prediction has long been viewed as essential to learning (e.g., Rescorla & Wagner 1972).

Prediction might play an important role in language development in several ways, such as when using transitional probabilities, when avoiding overgeneralizations, and when mapping form and meaning in novel phrasal constructions. Each of these three case studies is described, as follows.

***Transitional probabilities.*** Extracting the probability of Q given P can be useful in initial word segmentation (Graf Estes et al. 2007; Saffran et al. 1996), word learning (Hay et al. 2011; Mirman et al. 2008), and grammar learning (Gomez & Gerken 1999; Saffran 2002). A compelling way to interpret the contribution of transitional probabilities to learning is that P allows learners to form an expectation of Q (Turk-Browne et al. 2010). In fact, sensitivity to transitional probabilities correlates positively with the ability to use word predictability to facilitate comprehension under noisy input conditions (Conway et al. 2010). Moreover, sensitivity to sequential expectations also correlates positively with the ability to successfully process complex, long-distance dependencies in natural language (Misyak et al. 2010). Simple recurrent networks (SRNs) rely on prediction error to correct connection weights, and appear to learn certain aspects of language in much the same way as children do (Elman 1991; 1993; Lewis & Elman 2001; French et al. 2011).

***Statistical preemption.*** Children routinely make overgeneralization errors, producing *foots* instead of *feet*, or *She disappeared the quarter* instead of *She made the quarter disappear*. A number of theorists have suggested that learners implicitly predict upcoming formulations and compare witnessed formulations to their predictions, resulting in error-driven learning. That is, in contexts in which A is expected or predicted, but B is repeatedly used instead: children learn that B, not A, is the appropriate formulation – B *statistically preempts* A. Preemption is well accepted in morphology (e.g., *went* preempts *goed*; Aronoff 1976; Kiparsky 1982).

Unlike *went* and *goed*, distinct phrasal constructions are virtually never semantically and pragmatically identical. Nonetheless, if learners consistently witness one construction in contexts where they might have expected to hear another, the former can statistically preempt the latter (Goldberg 1995; 2006; 2011; Marcotte 2005). For example, if learners expect to hear *disappear* used transitively in relevant contexts (e.g., *She disappeared it*), but instead consistently hear it used periphrastically (e.g., *She made it disappear),* they appear to read just future predictions so that they ultimately prefer the periphrastic causative (Boyd & Goldberg 2011; Brooks & Tomasello 1999; Suttle & Goldberg forthcoming).

***Construction learning.*** Because possible sentences form an open-ended set, it is not sufficient to simply memorize utterances that have been heard. Rather, learners must generalize over utterances in order to understand and produce new formulations. The learning of novel phrasal constructions involves learning to associate form with meaning, such as the double object pattern with "intended transfer." Note, for example, that *She mooped him something* implies that she intends to *give* him something, and this meaning cannot be attributed to the nonsense word, *moop*. In the domain of phrasal construction learning, phrasal constructions appear to be at least as good predictors of overall sentence meaning as individual verbs (Bencini & Goldberg 2000; Goldberg et al. 2005).

We have recently investigated the brain systems involved in learning novel constructions. While undergoing functional magnetic resonance imaging (fMRI), participants were shown short audiovisual clips that provided the opportunity to learn novel constructions. For example, a novel "appearance" construction consisted of various characters appearing on or in another object, with the word order $Verb\text{-}NP_{\text{theme}}\text{-}NP_{\text{locative}}$, (where *NP* is noun phrase). For each construction, there was a patterned condition and a random condition. In the patterned condition, videos were consistently narrated by the $V\text{-}NP_{\text{theme}}\text{-}NP_{\text{locative}}$ pattern, enabling participants to associate the abstract form and meaning. In the random condition, the exact same videos were shown in the same order, but the words were randomly reordered; this inconsistency prevented successful construction learning. Most relevant to present purposes, we found an inverse relationship between ventral striatal (VS) activity and learning for patterned presentations only: Greater test accuracy on new instances (requiring generalization) was correlated with less ventral striatal activity during learning. In other tasks, VS gauges the discrepancy between predictions and outcomes, signaling that something new can be learned (Niv & Montague 2008; O'Doherty et al. 2004; Pagnoni et al. 2002). This activity may therefore suggest a role for prediction in construction learning: Better learning results in more accurate predictions of how the scene will unfold.

Such prediction-based learning may therefore be a natural consequence of making implicit predictions during language production and comprehension. Future research is needed to elucidate the scope of this prediction-based learning mechanism, and to understand its role in language. Such investigations would strengthen and ground P&G's proposal, and would suggest that predictions are central to both language use *and* language development.

## Communicative intentions can modulate the linguistic perception-action link

doi:10.1017/S0140525X12002610

Yoshihisa Kashima,[a] Harold Bekkering,[b] and Emiko S. Kashima[c]

[a]*Melbourne School of Psychological Sciences, The University of Melbourne, Parkville, Victoria 3010, Australia;* [b]*Donders Institute for Brain, Cognition, and Behaviour, Radboud University Nijmegen, 6500 HE Nijmegen, The Netherlands;* [c]*School of Psychological Science, La Trobe University, Bundoora, Victoria 3086, Australia.*

**ykashima@unimelb.edu.au** **h.bekkering@donders.ru.nl**
**e.kashima@latrobe.edu.au**
**http://www.psych.unimelb.edu.au/people/staff/KashimaY.html**
**http://www.nici.ru.nl/anc/index.php?staff=bekkering**
**http://www.latrobe.edu.au/scitecheng/about/staff/profile?uname=ekashima**

**Abstract:** Although applauding Pickering & Garrod's (P&G's) attempt to ground language use in the ideomotor perception-action link, which provides an "infrastructure" of embodied social interaction, we suggest that it needs to be complemented by an additional control mechanism