

# An integrated theory of language production and comprehension

**Martin J. Pickering**

*Department of Psychology, University of Edinburgh, Edinburgh EH8 9JZ, United Kingdom*

[martin.pickering@ed.ac.uk](mailto:martin.pickering@ed.ac.uk)

<http://www.ppls.ed.ac.uk/people/martin-pickering>

**Simon Garrod**

*University of Glasgow, Institute of Neuroscience and Psychology, Glasgow G12 8QT, United Kingdom*

[simon@psy.gla.ac.uk](mailto:simon@psy.gla.ac.uk)

<http://staff.psy.gla.ac.uk/~simon/>

**Abstract:** Currently, production and comprehension are regarded as quite distinct in accounts of language processing. In rejecting this dichotomy, we instead assert that producing and understanding are interwoven, and that this interweaving is what enables people to predict themselves and each other. We start by noting that production and comprehension are forms of action and action perception. We then consider the evidence for interweaving in action, action perception, and joint action, and explain such evidence in terms of prediction. Specifically, we assume that actors construct forward models of their actions before they execute those actions, and that perceivers of others' actions covertly imitate those actions, then construct forward models of those actions. We use these accounts of action, action perception, and joint action to develop accounts of production, comprehension, and interactive language. Importantly, they incorporate well-defined levels of linguistic representation (such as semantics, syntax, and phonology). We show (a) how speakers and comprehenders use covert imitation and forward modeling to make predictions at these levels of representation, (b) how they interweave production and comprehension processes, and (c) how they use these predictions to monitor the upcoming utterances. We show how these accounts explain a range of behavioral and neuroscientific data on language processing and discuss some of the implications of our proposal.

**Keywords:** comprehension; covert imitation; dialogue; forward model; language; prediction; production

## 1. Introduction

Current accounts of language processing treat production and comprehension as quite distinct from each other. The split is clearly reflected in the structure of recent handbooks and textbooks concerned with the psychology of language (e.g., Gaskell 2007; Harley 2008). This structure does not merely reflect organizational convenience but instead treats comprehension and production as two different questions to investigate. For example, researchers assume that the processes involved in comprehending a spoken or written sentence, such as resolving ambiguity, may be quite distinct from the processes involved in producing a description of a scene. In neurolinguistics, the “classic” Lichtheim–Broca–Wernicke model assumes distinct anatomical pathways associated with production and comprehension, primarily on the basis of deficit–lesion correlations in aphasia (see Ben Shalom & Poeppel 2008). This target article rejects such a dichotomy. In its place, we propose that producing and understanding are tightly interwoven, and this interweaving underlies people's ability to predict themselves and each other.

### 1.1. The traditional independence of production and comprehension

To see the effects of the split, we need to think about language use both within and between individuals, in terms of a model of communication (Fig. 1).

This model includes “thick” arrows between message and (linguistic) form, corresponding to production and comprehension. The production arrows represent the fact that production may involve converting one message into form (serial account) or the processor may convert multiple messages at once, then select one (parallel account). Within production, the “internal” arrows signify feedback (e.g., from phonology to syntax), which occurs in interactive accounts but not purely feedforward accounts. Note that these arrows are consistent with any type of information (linguistic or nonlinguistic) being used during production. The arrows play an analogous role within comprehension (e.g., the internal arrows could signify feedback from semantics to syntax). In contrast, the arrows corresponding to sound are “thin” because a single sequence of sounds is sent forward between the speakers. If communication is fully successful, then A's message<sub>1</sub>=B's message<sub>1</sub>. Similarly, there is a “thin” arrow for thinking because such accounts assume that each individual converts a single message (e.g., an understanding of a question, message<sub>1</sub>) into another (e.g., an answer, message<sub>2</sub>), and the answer does not affect the understanding of the question.

The model is split *vertically* between the processes in different individuals, who of course have independent minds. But it is also split *horizontally*, because the processes underlying production and comprehension within each individual are separated. The traditional model assumes discrete stages: one in which A is producing and

B is comprehending an utterance, and one in which B is producing and A is comprehending an utterance. Each speaker constructs a message that is translated into sound before the addressee responds with a new message. Hence, dialogue is “serial monologue,” in which interlocutors alternate between production and comprehension.

In conversation, however, interlocutors’ contributions often overlap, with the addressee providing verbal or non-verbal feedback to the speaker, and the speaker altering her contribution on the basis of this feedback. In fact, such feedback can dramatically affect both the quality of the speaker’s contribution (e.g., Bavelas et al. 2000) and the addressee’s understanding (Schober & Clark 1989). This of course means that both interlocutors must simultaneously produce their own contributions and comprehend the other’s contribution. Clearly, an approach to language processing that assumes a temporal separation between production and comprehension cannot explain such behavior.

Interlocutors are not static, as the traditional model assumes, but are “moving targets” performing a joint activity (Garrod & Pickering 2009). They do not simply transmit messages to each other in turn but rather negotiate the form and meaning of expressions they use by interweaving their contributions (Clark 1996), as illustrated in (1a–1c), below (from Gregoromichelaki et al. 2011). In (1b), B begins to ask a question, but A’s interruption (1c) completes the question and answers it. B, therefore, does not discretely encode a complete message into sound but, rather, B and A jointly encode the message across (1b–c).

1a—A: I’m afraid I burnt the kitchen ceiling

1b—B: But have you

1c—A: burned myself? Fortunately not.

MARTIN J. PICKERING is Professor of the Psychology of Language and Communication at the University of Edinburgh. He is the author of more than 100 journal articles and numerous other publications in language production, comprehension, dialogue, reading, and bilingualism. His articles have appeared in *Psychological Bulletin*; *Trends in Cognitive Sciences*; *Psychological Science*; *Cognitive Psychology*; *Journal of Memory and Language*; *Cognition*; *Journal of Experimental Psychology: Learning, Memory, and Cognition*; and many other journals. In particular, he published “Toward a mechanistic psychology of dialogue” (*Behavioral and Brain Sciences*, 2004) with Simon Garrod. He is a Fellow of the Royal Society of Edinburgh and the editor of *Journal of Memory and Language*.

SIMON GARROD is Professor of Cognitive Psychology at the University of Glasgow. Between 1989 and 1999 he was also Deputy Director of the ESRC Human Communication Research Centre. He has published two books, one with Anthony J. Sanford, *Understanding written language*, and one with Kenny R. Coventry, *Seeing, saying and acting: The psychological semantics of spatial prepositions*. Additionally, he has published more than 100 research papers on various aspects of the psychology of language. His special interests include discourse processing, language processing in dialogue, psychological semantics, and graphical communication. He is a Fellow of the Royal Society of Edinburgh.

The horizontal split is also challenged by findings from isolated instances of comprehension or production. Take picture-word interference, in which participants are told to name a picture (e.g., of a dog) while ignoring a spoken or written distractor word (e.g., Schriefers et al. 1990). At certain timings, they are faster naming the picture if the word is phonologically related to it (*dot*) than if it is not. The effect cannot be caused by the speaker’s interpreting *dot* before producing *dog*—the meaning of *dot* is not the cause of the facilitation. Rather, the participant accesses phonology during the comprehension of *dot*, and this affects the construction of phonology during the production of *dog*. So experiments such as these suggest that production and comprehension are tightly interwoven. Quite ironically, most psycholinguistic theories attempt to explain either production *or* comprehension, but a great many experiments appear to involve both. Single word naming is typically used to explain comprehension but involves production (see Bock 1996). Sentence completion is often used to explain production but involves comprehension (e.g., Bock & Miller 1991). Similarly, the finding that word identification can be affected by externally controlled cheek movement (Ito et al. 2009) suggests that production influences comprehension.

In addition, production and comprehension appear to recruit strongly overlapping neural circuits (Scott & Johnsrude 2003; Wilson et al. 2004). For example, Paus et al. (1996) found activation (dependent on the rate of speech) of regions associated with speech perception when people whispered but could not hear their own speech. Listeners also activate appropriate muscles in the tongue and lips while listening to speech but not nonspeech (Fadiga et al. 2002; Watkins et al. 2003). Additionally, increased muscle activity in the lips is associated with increased activity (i.e., blood flow) in Broca’s area, suggesting that this area mediates between the comprehension and production systems during speech perception (Watkins & Paus 2004). There is also activation of brain areas associated with production during aspects of comprehension from phonology (Heim et al. 2003) to narrative structure (Mar 2004; see Scott et al. (2009) and Pulvermüller and Fadiga (2010). Finally, Menenti et al. (2011) found massive overlap between speaking and listening for regions showing functional magnetic resonance imaging (fMRI) adaptation effects associated with repeating language at different linguistic levels (see also Segaert et al. 2012). These results are inconsistent with separation of neural pathways for production and comprehension in the classical Lichtheim–Broca–Wernicke neurolinguistic model.

In conclusion, the evidence from dialogue, psycholinguistics, and cognitive neuroscience all casts doubt on the independence of production and comprehension, and therefore on the horizontal split assumed in Figure 1. Let us now address two theoretical issues relating to the abandonment of this split, and then ask what kind of model is compatible with the interweaving of production and comprehension.

## 1.2. Modularity and the cognitive sandwich

Much of psycholinguistics has sought to test the claim that language processing is modular (Fodor 1983). Such accounts investigate the way in which information travels

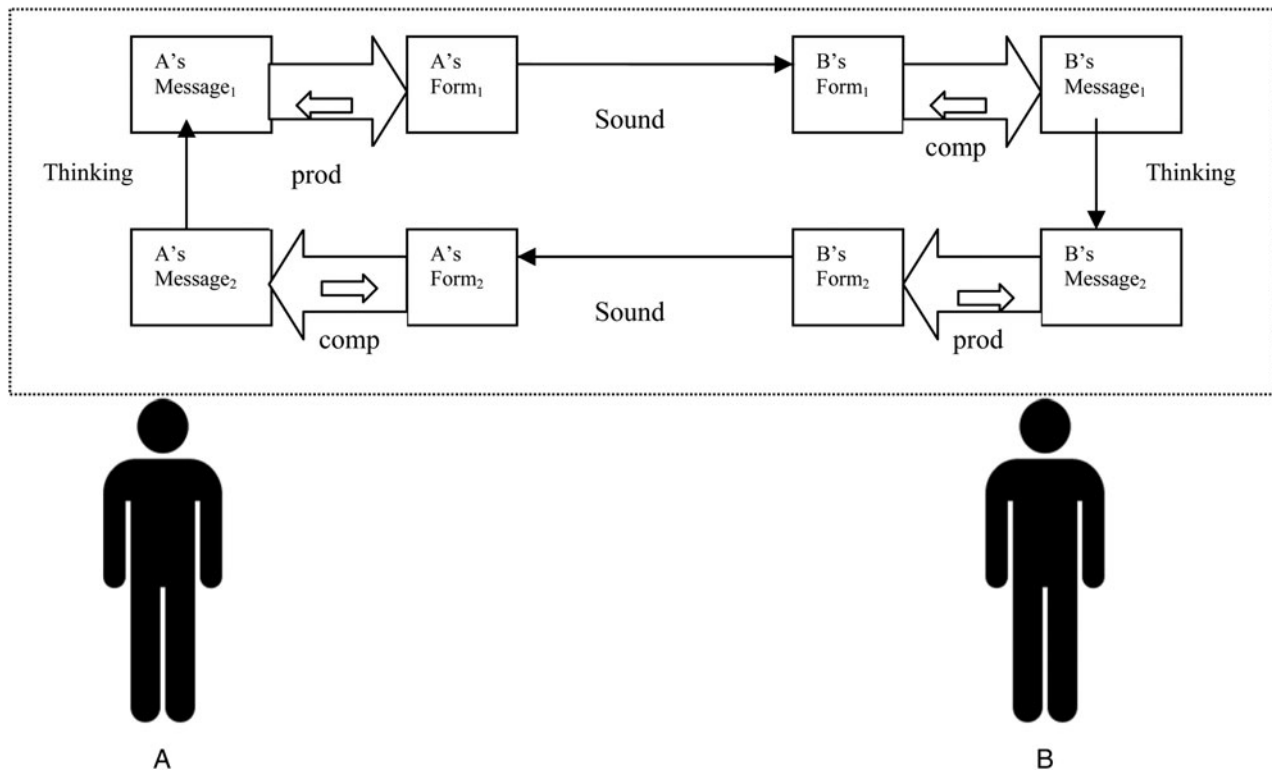


Figure 1. A traditional model of communication between A and B. (comp: comprehension; prod: production)

between the boxes in a model such as in Figure 1. In particular, the arrows labeled *thinking* correspond to “central processes” and contain representations in some kind of language of thought. Researchers are particularly concerned with the extent to which *thinking* arrows are separated from the *production* and *comprehension* arrows. Modular theories assume that some aspects of production or comprehension do not make reference to “central processes” (e.g., Frazier 1987; Levelt et al. 1999). In contrast, interactionist theories allow “central processes” to directly affect production or comprehension (e.g., Dell 1986; MacDonald et al. 1994; Trueswell et al. 1994). But both types of theory maintain that production and comprehension are separated from each other. In this sense, both types of theory are modular and are compatible with Figure 1.

In fact, Hurley (2008a) argued that traditional cognitive psychology assumes this type of modularity in order to keep action and perception separate. She referred to this assumption as the *cognitive sandwich*. Individuals perceive the world, reason about their perceptions using thinking (i.e., cognition), and act on the basis of those thoughts. Researchers assume that action and perception involve separate representations and processes and study one or the other but not both (and they are kept separate in textbooks and the like). In Hurley’s terms, the cognitive “meat” keeps the motor “bread” separate from the perceptual “bread.”<sup>1</sup> She argued that perception and action are interwoven and, therefore, rejected the cognitive sandwich.

Importantly, language production is a form of action and language comprehension is a form of perception. Therefore, traditional psycholinguistics also assumes the cognitive sandwich, with the thinking “meat” keeping apart the production and comprehension “bread.” But if action and perception are interwoven, then production and comprehension are

interwoven as well, and so accounts of language processing should also reject the cognitive sandwich.

### 1.3. Production and comprehension processes

How can production and comprehension both be involved in isolated speaking or listening? Within the individual, we mean that production and comprehension *processes* are interwoven. Production processes must of course be used when individuals produce language, and comprehension processes must be used when they comprehend language. However, production processes must also be used during, for example, silent naming, when no utterance is produced. Silent naming therefore involves some production processes (e.g., those associated with aspects of formulation such as name retrieval) but not others (e.g., those associated with articulation; see Levelt 1989). Likewise, comprehension processes must occur when a participant retrieves the phonology of a masked prime word but not its semantics (e.g., Van den Bussche et al. 2009). And so it is also possible that production processes are used during comprehension and comprehension processes used during production.

How can we distinguish production processes from comprehension processes? For this, we assume that (1) people represent linguistic information at different levels; (2) these levels are semantics, syntax, and phonology<sup>2</sup>; (3) they are ordered “higher” to “lower,” so that a speaker’s message is linked to semantics, semantics to syntax, syntax to phonology, and phonology to speech. We then assume that a producer goes from message to sound via each of these levels (message → semantics → syntax → phonology → sound), and a comprehender goes from sound to message in the opposite direction. Given this framework, we

define a *production process* as a process that maps from a “higher” to a “lower” linguistic level (e.g., syntax to phonology) and a *comprehension process* as a process that maps from a “lower” to a “higher” level.<sup>3</sup> This means that producing utterances must involve production processes, but can also involve comprehension processes; similarly, comprehending utterances must involve comprehension processes, but can also involve production processes.

One possibility is that people have separate production and comprehension systems. On this account, producing utterances may make use of feedback mechanisms that are similar in some respects to the mechanisms of comprehension, and comprehending utterances may make use of feedback mechanisms that are similar in some respects to the mechanisms of production. This is the position assumed by traditional interactive models of production (e.g., Dell 1986) and comprehension (e.g., MacDonald et al. 1994). In such accounts, production and comprehension are internally nonmodular, but are modular with respect to each other. They do not take advantage of the comprehension system in production or the production system in comprehension (even though the other system is often lying dormant).

Very little work in comprehension makes reference to production processes, with classic theories of lexical processing (from, e.g., Marslen-Wilson & Welsh 1978 or Swinney 1979 onward) and sentence processing (e.g., Frazier 1987; MacDonald et al. 1994) making no reference to production processes (see Bock 1996 for discussion, and Federmeier 2007 for an exception). In contrast, some theories of production do incorporate comprehension processes. Most notably, Levelt (1989) assumed that speakers monitor their own speech using comprehension processes. They can hear their own speech (external self-monitoring), in which case the speaker comprehends his own utterance just like another person’s utterance; but they can also monitor a sound-based representation (internal self-monitoring), in which comprehension processes are used to convert sound to message (see sect. 3.1).

In addition, some computationally sophisticated models can use production and comprehension processes together (e.g., Chang et al. 2006), use comprehension to assist in the process of learning to speak (Plaut & Kello 1999), or assume that comprehension and production use the same network of nodes and connections so that feedback processes during production are the same as feedforward processes during comprehension (MacKay 1982). In addition, Dell has proposed accounts in which feedback during production is a component of comprehension (e.g., Dell 1988), although he has also queried this claim on the basis of neuropsychological evidence (Dell et al. 1997, p. 830); see also the debate between Rapp and Goldrick (2000; Rapp & Goldrick 2004) and Roelofs (2004).

But none of these theories incorporate mechanisms of sentence comprehension (e.g., parsing or lexical ambiguity resolution) into theories of production. We believe that this is a consequence of the traditional separation of production and comprehension (as represented in Fig. 1). In contrast, we propose that comprehension processes are routinely accessed at different stages in production, and that production processes are routinely accessed at different stages in comprehension.

The rest of this target article develops an account of language processing in which processes of production and

comprehension are integrated. We assume that instances of both production and comprehension involve extensive use of prediction – determining what you yourself or your interlocutor is likely to say next. Predicting your own utterance involves comprehension processes as well as production processes, and predicting another person’s utterance involves production processes as well as comprehension processes.

As we have noted, production is a form of action, and comprehension is a form of perception. More specifically, comprehension is a form of *action perception* – perception of other people performing actions. We first consider the evidence for interweaving in action and action perception, and we explain such evidence in terms of prediction. We assume that actors construct forward models of their actions before they execute those actions, and that perceivers of others’ actions construct forward models of others’ actions that are based on their own potential actions. Finally, we apply these accounts to joint action.

We then develop these accounts of action, action perception, and joint action into accounts of production, comprehension, and dialogue. Unlike many other forms of action and perception, language processing is clearly structured, incorporating well-defined levels of linguistic representation such as semantics, syntax, and phonology. Thus, our accounts also include such structure. We show how speakers and comprehenders predict the content of levels of representation by interweaving production and comprehension processes. We then explain a range of behavioral and neuroscientific data on language processing, and discuss some of the implications of the account.

## 2. Interweaving in action and action perception

For perception and action to be interwoven, there must be a direct link between them. If so, there should be much evidence for effects of perception on action, and there is. In one study, participants’ arm movements showed more variance when they observed another person making a different versus the same arm movement (Kilner et al. 2003; see also Stanley et al. 2007). Conversely, there is good evidence for effects of action on perception. For example, producing hand movements can facilitate the concurrent visual discrimination of deviant hand postures (Miall et al. 2006), and turning a knob can affect the perceived motion of a perceptually bistable object (Wohlschläger 2000). Such evidence immediately casts doubt on the “sandwich” architecture for perception and action.

What purpose might such a link serve? First, it could facilitate overt imitation, but overt imitation is not common in many species (see Prinz 2006). Second, it could be used *postdictively*, with action representations helping perceivers develop a stable memory for a percept or a detailed understanding of it (e.g., via rehearsal), and perceptual representations doing the same for actors. But we propose a third alternative: people compute action representations during perception and perception representations during action to aid *prediction* of what they are about to perceive or to do, in a way that allows them to “get ahead of the game” (see Wilson & Knoblich 2005).<sup>4</sup> To explain this, we turn to the theory of forward modeling, which was first applied to action but has more recently been applied to action perception. We interpret the theory in a

way that then allows us to extend it to account for language processing.

## 2.1. Forward modeling in action

To explain forward modeling, we draw on Wolpert's proposals from computational neuroscience (e.g., Davidson & Wolpert 2005; Wolpert 1997), but reframed using psychological terminology couched in the language of perception and action (see Fig. 2). We use the simple example of moving a hand to a target. The actor formulates the action (motor) command to move the hand. This command initiates two processes in parallel. First, it causes the action implementer to generate the act, which in turn leads the perceptual implementer to construct a percept of the experience of moving the hand. In Wolpert's terms, this percept is used as sensory feedback (*reaffERENCE*) and is partly proprioceptive, but may also be partly visual (if the agent watches her hand move).

Second, it sends an *effERENCE COPY* of the action command to cause the forward action model to generate the predicted act of moving the hand.<sup>5</sup> Just as the act depends on the application of the action command to the current state of the action implementer (e.g., where the hand is positioned before the command), so the predicted act depends on the application of the effERENCE COPY of the action command to the current state of the forward action model (e.g., a model of where the hand is positioned before

the command). The predicted act then causes the forward perceptual model to construct a predicted percept of the experience of moving the hand. (This percept would not form part of a traditional action plan.) Note that this predicted percept is compatible with the theory of event coding (Hommel et al. 2001), in which actions are represented in terms of their predicted perceptual consequences.

Importantly, the effERENCE COPY is (in general) processed more quickly than the action command itself (see Davidson & Wolpert 2005). For example, the command to move the hand causes the action implementer to activate muscles, which is a comparatively slow process. In contrast, the forward action model and the forward perceptual model make use of representations of the position of the hand, state of the muscles, and so on (and may involve simplifications and approximations). These representations may be in terms of equations (e.g., hand coordinates), and such equations can (typically) be solved rapidly (e.g., using a network that represents relevant aspects of mathematics). So the predicted percept (the predicted sensations of the hand's movement and position) is usually "ready" before the actual percept. The action then occurs and the predicted percept is compared to the actual percept (the sensations of the hand's actual movement and position).

Any discrepancy between these two sensations (as determined by the comparator) is fed back so that it can modify the next action command accordingly. If the hand is to the

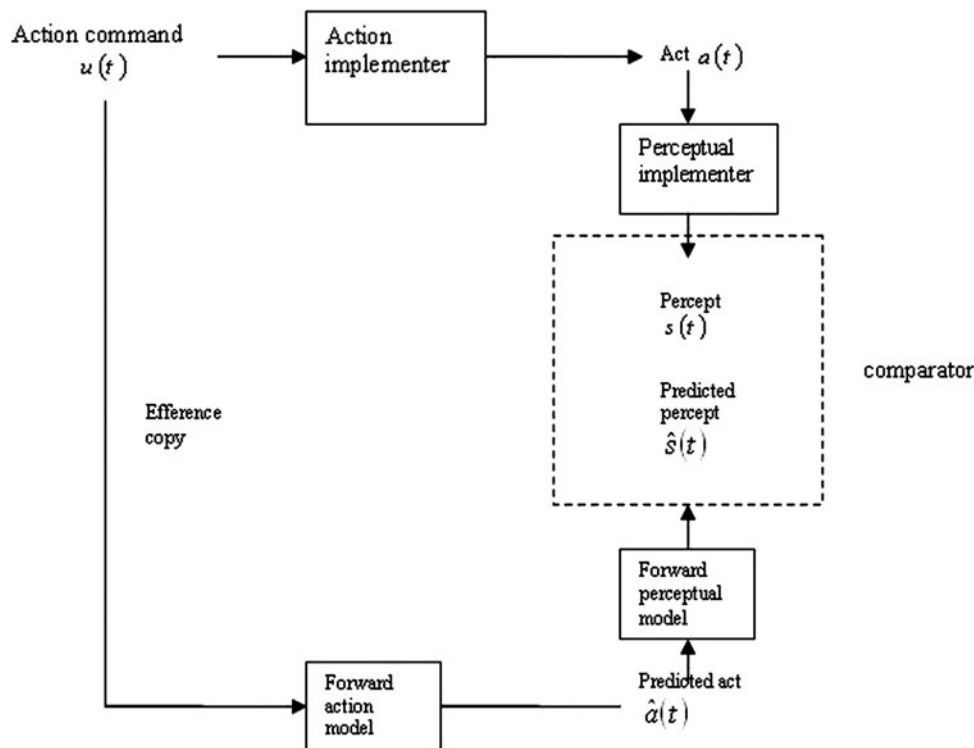


Figure 2. A model of the action system, using a snapshot of executing an act at time  $t$ . Boxes refer to processes, and terms not in boxes refer to representations. The action command  $u(t)$  (e.g., to move the hand) initiates two processes. First,  $u(t)$  feeds into the action (motor) implementer, which outputs an act  $a(t)$  (the event of moving the hand). In turn, this act feeds into the perceptual (sensory) implementer, which outputs a percept  $s(t)$  (the perception of moving the hand). Second, an effERENCE COPY of  $u(t)$  feeds into the forward action model, a computational device (distinct from the action implementer) which outputs a predicted act  $\hat{a}(t)$  (the predicted event of moving the hand); the carat indicates an approximation. In turn,  $\hat{a}(t)$  feeds into the forward perceptual model, a computational device (distinct from the perceptual implementer) which outputs a predicted percept  $\hat{s}(t)$  (the predicted perception of moving the hand). The comparator can be used to compare the percept and the predicted percept.

left of its predicted position, the next action command can move it more to the right. In this way, perceptual processes have an online effect on action, so that the act can be repeatedly affected by perceptual processes as well as action processes. (Alternatively, the actor can correct the forward model rather than the action command, depending on her confidence about the relative accuracy of the action command and the efference copy.) Such prediction is necessary because determining the discrepancy on the basis of reafferent feedback would be far too slow to allow corrective movements (see Grush [2004], who referred to forward models as *emulators*).

We assume that the central role of forward modeling is perceptual prediction (i.e., predicting the perceptual outcomes of an action). However, it has other functions. First, it can be used to help estimate the current state, given that perception is not entirely accurate. The best estimate of the current position of the hand combines the estimate that comes from the percept and the estimate that comes from the predicted percept. For example, a person can estimate the position of her hand in a dark room by remembering the action command that underlay her hand movement to its current location. Second, forward models can cancel the sensory effects of self-motion (*reafference cancellation*) when these sensory effects match the predicted movement. This enables people to differentiate between perceptual effects of their own actions and those reflecting changes in the world, for example, explaining why self-applied tickling is not effective (Blakemore et al. 1999).

A helpful analogy is that of an old-fashioned sailor navigating across the ocean (cf. Grush 2004). He starts at a known position, which he marks on his chart (i.e., model of the ocean), and determines a compass course and speed. He lays out the corresponding course on the chart and traces out where he should be at noon (his predicted act,  $\hat{a}(t)$ ), and determines what his sextant should read at this time and place (his predicted percept,  $\hat{s}(t)$ ). He then sets off through the water until noon (his act,  $a(t)$ ). At noon, he uses his sextant to estimate his position from the sun (his percept,  $s(t)$ ), and compares the predicted and observed sextant readings (using the comparator). He can then use this in various ways. If he is not confident of his course keeping, he pays more attention to the actual reading; if he is not confident of his sextant reading (e.g., it is misty), he pays more attention to the predicted reading. If the predicted and actual readings match, he assumes no other force (this is equivalent to reafference cancellation). But if they do not match and he is confident about both course keeping and sextant reading, he assumes the existence of another force, in this case the current.

Forward modeling also plays an important role in motor learning (Wolpert 1997). To be able to pick up an object you need a model that maps the object's location onto an action (motor) command to move the hand to that location. This is called an inverse model because it represents the inverse of the forward model. Learning a motor skill requires learning both an appropriate forward model and an appropriate inverse model.

Motor control theories that are more sophisticated use linked forward-inverse model pairs to explain how actors can adapt dynamically to changes in the context of an unfolding action. In their Modular Selection and Identification for Control (MOSAIC) account, Haruno et al.

(2001) proposed that actors run sets of model pairs in parallel, with each forward model making different predictions about how the action might unfold in different contexts. By matching actual movements against these different predictions, the system can shift responsibility for controlling the action toward the model pair whose forward model prediction best fits that movement. For example, a person starts to pick up a small (and apparently light) object using a weak grip but subsequently finds the grip insufficient to lift the object. According to MOSAIC, the person would then shift the responsibility for controlling the action to a new forward-inverse model pairing, which produces a stronger grip.

The same principles apply to structured activities that are more complex, such as the process of drinking a cup of tea. Here the forward model provides information ahead of time about the sequence and overlap between the different stages in the process (moving the hand to the cup, picking it up, moving it to the mouth, opening the mouth, etc.) and represents the predicted sensory feedback at each stage (i.e., the predicted percept). Controlling such complex sequences of actions has been implemented by Haruno et al. (2003) in their Hierarchical MOSAIC (HMOSAIC) model. HMOSAIC extends MOSAIC by having hierarchically organized forward-inverse model pairings that link "high level" intentions to "low level" motor operations – in our terms, from high-level to low-level action commands.

In conclusion, forward modeling in action allows the actor to predict her upcoming action, in a way that allows her to modify the unfolding action if it fails to match the prediction. In addition, it can be used to facilitate estimation of the current state, to cancel reafference, and to support short- and long-term learning. In doing so, forward modeling closely interweaves representations associated with action and representations associated with perception, and can therefore explain effects of perception on action.

## 2.2. Covert imitation and forward modeling in action perception

When you perceive inanimate objects, you draw on your perceptual experience of objects. For example, if an object's movement is unclear, you can think about how similar objects have appeared to move in the past (e.g., obeying gravity). When you perceive other people (i.e., *action perception*), you can also draw on your perceptual experience of other people. We refer to this as *the association route* in action perception. For example, you assume someone's ambiguous arm movement is compatible with your experience of perceiving other people's arm movements. People can clearly predict each other's actions using the association route, just as they can predict the movement of physical objects on the basis of past experience (e.g., Freyd & Finke 1984).

However, you can also draw on your experience of your own body—you assume that someone's arm movement is compatible with your experience of your own arm movements. We refer to this as the *simulation route* in action perception. The simplest possibility is that the perceiver determines what she would do under the circumstances. In the case of hand movement, the perceiver would see the start of the actor's hand movement and would then determine how she would move if it were her hand,

thereby determining the actor's intention. Informally, she would see the hand and the way it was moving, and then think of it as her own hand and use the mechanisms that she would use to move her own hand to predict her partner's hand movement. In other words, she would covertly imitate her partner's movements, treating his arm positions as though they were her own arm positions. However, the perceiver cannot simply use the same mechanisms the actor would use but must "accommodate" to the differences in their bodies (the *context*, in motor control theory) – for example, applying a smaller force if her body is lighter weight than her partner's.<sup>6</sup> In any case, her reproduction is unlikely to be perfect – she is in the position of a character actor attempting to reproduce another person's mannerisms.

In theory, the perceiver could simulate by using her own action implementer (and inhibiting its output). However, this would be too slow – much of the time, she would determine her partner's action after he had performed that action. Instead, she can use her forward action model to derive a prediction of her partner's act (and the forward

perceptual model to derive a prediction of her percept of that act). To do this, she would identify the actor's intention from her perception of the previous and current states of his arm (or from background information such as knowledge of his state of mind) and use this to generate an efference copy of the intended act. If she determined that the actor was about to punch her face, she would have time to move. She can also compare this predicted percept with her actual percept of his act when it happens. We illustrate this account in Figure 3.

This simulation account uses the mechanisms involved in the prediction of action (as illustrated in Fig. 2), but it adds a mechanism for covert imitation. This mechanism also allows for overt imitation of the action itself or a continuation of that action (the overt imitation and continuation are what we call *overt responses*). In fact, the strong link between actions and predictions of those actions means that perceivers tend to activate their action implementers as well as forward action models. Note that Figure 3 ignores the association route to action prediction, which uses the percept  $s_B(t)$  and knowledge about percepts that

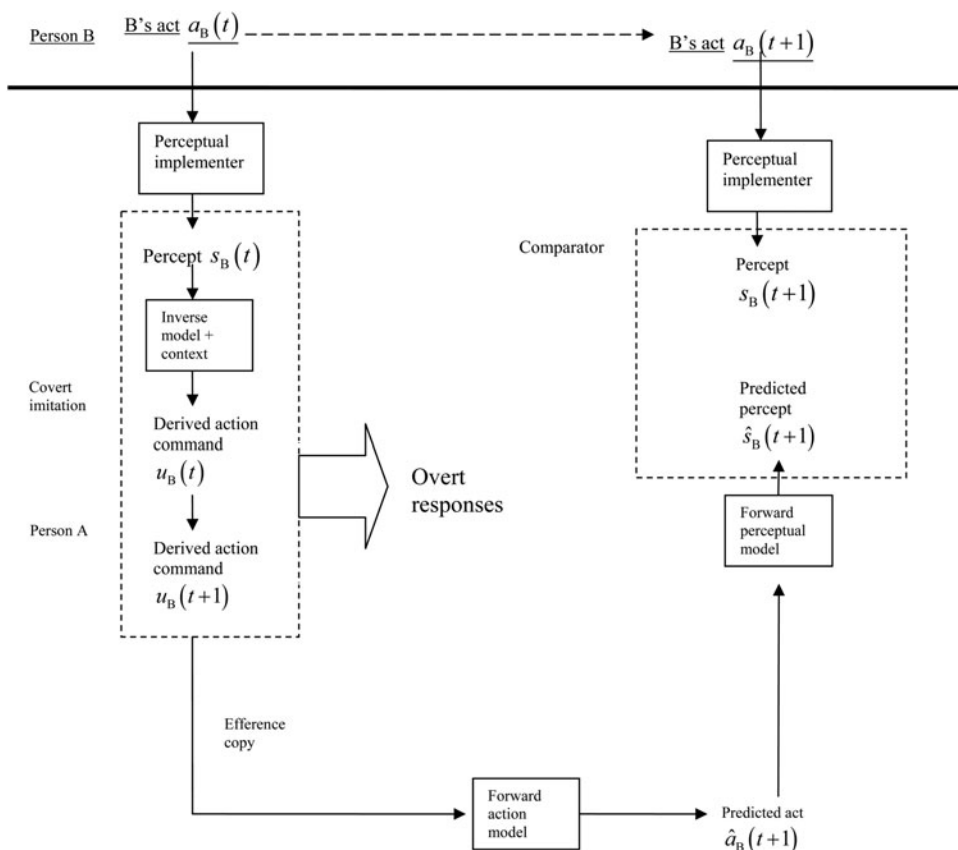


Figure 3. A model of the simulation route to prediction in action perception in Person A. Everything above the solid line refers to the unfolding action of Person B (who is being observed by A), and we underline B's representations. For instance,  $a_B(t)$  can refer to B's initial hand movement (at time  $t$ ) and  $a_B(t+1)$  to B's final hand movement (at time  $t+1$ ). A predicts B's act  $a_B(t+1)$  given B's act  $a_B(t)$ . To do this, A first covertly imitates B's act. This involves perceiving B's act  $a_B(t)$  to derive the percept  $s_B(t)$ , and from this using the inverse model and context (e.g., information about differences between A's body and B's body) to derive the action command (i.e., the intention)  $u_B(t)$  that A would use if A were to perform B's act (without context, the inverse model would derive the command that B would use to perform B's act—but this command is useless to A) and from this the action command that A would use if A were to perform the subsequent part of B's act  $u_B(t+1)$ . A now uses the same forward modeling that she uses when producing an act (see Fig. 2) to produce her prediction of B's act  $\hat{a}_B(t+1)$ , and her prediction of her perception of B's act  $\hat{s}_B(t+1)$ . This prediction is generally ready before her perception of B's act  $s_B(t+1)$ . She can then compare  $\hat{s}_B(t+1)$  and  $s_B(t+1)$  using the comparator. Notice that A can also use the derived action command  $u_B(t)$  to overtly imitate B's act and the derived action command  $u_B(t+1)$  to overtly produce the subsequent part of B's act (see "Overt responses").

tend to follow  $s_B(t)$  to predict the percept of the act. (The perceiver may of course be able to combine the action-based and perceptual predictions into a single prediction.) Additionally, we have glossed over the computationally complex part of this proposal—the mapping from the percept  $s_B(t)$  to the action commands  $u_B(t)$  and  $u_B(t+1)$ . How can the perceiver determine the actor's intention?

In fact, Wolpert et al. (2003) showed how to do this using HMOSAIC, which can make predictions about how different intentional acts unfold over time. In their account, the perceiver runs parallel, linked forward-inverse model pairings at multiple levels from “low-level” movements to “high-level” intentions. By matching actual movements against these different predictions, HMOSAIC determines the likelihood of different possible intentions (and dynamically modifies the space of possible intentions). This in turn modifies the perceiver's predictions of the actor's likely behavior. For example, a first level might determine that a movement of the shoulder is likely to lead to a movement of the arm (and would draw on information about the actor's body shape); a second level might determine whether such an arm movement is the prelude to a proffered handshake or a punch (and would draw on information about the actor's state of mind). At the second level, the perceiver runs forward models based on those alternative intentions to determine what the actor's hand is likely to do next. If, for example, I predict you are more likely to initiate a handshake but then your fist starts clenching, I modify my interpretation of your intention and now predict that you will likely throw a punch. At this point, I have determined your intention and confidently predict the upcoming position of your hand, just as I would do if I were predicting my own hand movements.

Good evidence that covert imitation plays a role in prediction comes from studies showing that appropriate motor-related brain areas can be activated before a perceived event occurs (Haueisen & Knösche 2001). Similarly, mirror neurons in monkeys can be activated by perceptual predictions as well as by perceived actions (Umiltà et al. 2001); note there is recent direct evidence for mirror neurons in people (Mukamel et al. 2010).<sup>7</sup> Additionally, people are better at predicting a movement trajectory (e.g., in dart-throwing or handwriting) when viewing a video of themselves versus others (Knoblich & Flach 2001; Knoblich et al. 2002). Presumably, prediction-by-simulation is more accurate when the object of the prediction is one's own actions than when it is someone else's actions. This yoking of action-based and perceptual processes can therefore explain the experimental evidence for interweaving (e.g., Kilner et al. 2003).

Notice that such covert imitation can also drive overt imitation. However, the perceiver does not simply copy the actor's movements, but rather bases her actions on her determination of the actor's intentions. This is apparent in infants' imitation of caregivers' actions (Gergely et al. 2002) and in the behavior of mirror neurons, which code for intentional actions (Umiltà et al. 2001). Importantly, mirror neurons do not exist merely to facilitate imitation (because imitation is largely or entirely absent in monkeys), and so one of their functions may be to drive action prediction via covert imitation (Csibra & Gergely 2007; Prinz 2006). In conclusion, we propose that action perception interweaves action-based and perceptual processes in a way that supports prediction.

### 2.3. Joint action

People are highly adept at joint activities, such as ballroom dancing, playing a duet, or carrying a large object together (Sebanz et al. 2006a). Clearly, such activities require two (or more) agents to coordinate their actions, which in turn means that they are able to perceive each other's acts and perform their own acts together. In many of these activities, precise timing is crucial, with success occurring only if each partner applies the right force at the right time in relation to the other. Such success therefore requires tight interweaving of perception and action. Moreover, people must predict each other's actions, because responding after they perceive actions would simply be too slow. Clearly, it may also be useful to predict one's own actions, and to integrate these predictions with predictions of others' actions.

We therefore propose that people perform joint actions by combining the models of prediction in action and action perception in Figures 2 and 3. Figure 4 shows how A and B can both predict B's upcoming action (using prediction-by-simulation). A perceives B's current act and then uses covert imitation and forward modeling; B formulates his forthcoming act and uses forward modeling based on that intention. If successful, A and B should make similar predictions about B's upcoming act, and they can use those predictions to coordinate. Note that they can both compare their predictions with B's forthcoming act when it takes place.

Joint action can involve overt imitation, continuation of other's behavior, or complementary behavior. Overt imitation and continuation follow straightforwardly from Figure 3 (see the large arrow leading from “Covert imitation” to “Overt responses”). There is much evidence that people overtly imitate each other without intending to or being aware that they are doing so, from studies involving

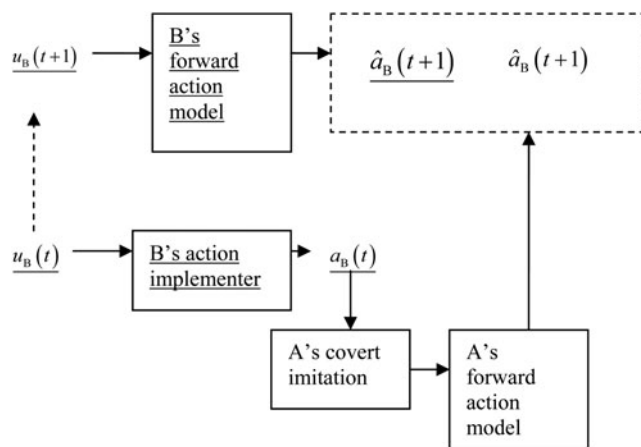


Figure 4. A and B predicting B's forthcoming action (with B's processes and representations underlined). B's action command  $u_B(t)$  feeds into B's action implementer and leads to B's act  $a_B(t)$ . A covertly imitates B's act and uses A's forward action model to predict B's forthcoming act (at time  $t+1$ ). B simultaneously generates the next action command (the dotted line indicates that this command is causally linked to the previous action command for B but not A) and uses B's forward action model to predict B's forthcoming act. If A and B are coordinated, then A's prediction of B's act and B's prediction of B's act (in the dotted box) should match. Moreover, they may both match B's forthcoming act at time  $t+1$  (not shown). A and B also predict A's forthcoming action (see text).

the imitation of specific movements (e.g., Chartrand & Bargh, 1999; Lakin & Chartrand, 2003) or involving the synchronization of body posture (e.g., Shockley et al., 2003). For example, pairs of participants tend to start rocking chairs at the same frequency, even though the chairs have different natural frequencies (Richardson et al., 2007), and crowds come to clap in unison (Neda et al., 2000). Such imitation appears to be on a perception-behavior expressway (Dijksterhuis & Bargh, 2001), not mediated by inference or intention. Many of these findings demonstrate close temporal coordination and appear to require prediction (see Sebanz & Knoblich, 2009). For instance, in a joint go/no-go task, Sebanz et al. (2006b) found enhanced N170 event-related potentials (ERPs), reflecting response inhibition, for the nonresponding player when it was the partner's turn to respond. They interpreted this as suggesting that a person suppresses his or her own actions at the point when a partner is about to act. In addition, people continue each other's behavior by overtly imitating their predicted behavior (in contrast to overt imitation of actual behavior). For example, early studies showed that some mirror neurons fired when the monkey observed a matching action (i.e., one that would cause that neuron to fire if the monkey performed that action) and others fired when it observed a nonmatching action that could precede the matching action (di Pellegrino et al., 1992).

Complementary behavior occurs when the co-actors use their same predictions to derive different (but coordinated) behaviors. For example, in ballroom dancing, both A and B predict that B will move his foot forward; B will then move his foot, and A will plan her complementary action of moving her foot backward. Graf et al. (2010) reviewed much evidence for complementary motor involvement in action perception (see Häberle et al. 2008; Newman-Norlund et al. 2007; van Schie et al. 2008).

So far, we have described how A and B predict B's action. To explain joint activity, we first note that A and B predict A's action as well (in the same way). They then integrate these predictions with their predictions of B's action. To do this, they must simultaneously predict their own action and their partner's action. They can determine whether these acts are compatible (essentially asking themselves, "does my upcoming act fit with your upcoming act?"). If not, they can modify their own upcoming actions accordingly (so that such modifications can occur on the basis of comparing predictions alone, without having to wait for the action). (If I find out that I am likely to collide with you, I can move out of the way.) This account can therefore explain tight coupling of joint activity, as well as the experience of "shared reality" that occurs when A and B realize that they are experiencing the world in similar ways (Echterhoff et al. 2009).

Importantly, the participants in a joint action perform actions that are related to each other. It is of course easier for A to predict both A and B's actions if their actions are closely related (as is the case in tightly coupled activities such as ballroom dancing). If A's predictions of her own action ( $\hat{a}_A(t+1)$ ) and her prediction of B's action ( $\hat{a}_B(t+1)$ ) were unrelated, she would find both predictions hard; if the predictions are closely related, A is able to use many of the computations involved in one prediction to support the other prediction. In other words, it is easier to predict another person's actions when you are

performing a related action than when you are performing an unrelated action. (Notice also that A and B are likely to overtly imitate each other and that such overt imitation will make their actions more similar, hence the predictions easier to integrate.) In conclusion, joint action can be successful because the participants are able to integrate their own action with their perception of their partner's action.

### 3. A unified framework for language production and comprehension

We have noted that language production is a form of action and comprehension is a form of action perception; accordingly, we now apply the above framework to language. This is of course consistent with the evidence for interweaving that we briefly considered in section 1: the tight coupling between interlocutors in dialogue, the evidence for effects of comprehension processes on acts of production and vice versa in behavioral experiments, and the overlap of brain circuits involved in acts of production and comprehension. We now argue that such interweaving occurs primarily to facilitate prediction, which in turn facilitates production and comprehension.

We first propose that speakers use forward production models of their utterances in the same way that actors use forward action models, by constructing efference copies of their predicted utterance and comparing those copies with the output of the production implementer. We then propose that listeners predict speakers' upcoming utterances by covertly imitating what they have uttered so far, deriving their underlying message, generating efference copies, and comparing those copies with the actual utterances when they occur, just as in our account of action perception. Dialogue involves the integration of the models of the speaker and the listener. These proposals are directly analogous to our proposals for action, action perception, and joint action, except that we assume structured representations of language involving (at least) semantics, syntax, and phonology.

#### 3.1. Forward modeling in language production

In acting, the action command drives the action implementer to produce an act, which the perceptual implementer uses to produce a percept of that act (see Fig. 2). But typically, before this process is complete, the efference copy of the action command drives the forward action model to produce a predicted act, which the forward perceptual model uses to produce a predicted percept. The actor can then compare these outputs and adjust the action command (or the forward model) if they do not match.

In language production (see Fig. 5), the action command is specified as a production command. The action implementer is specified as the production implementer, and the perceptual implementer is specified as the comprehension implementer. Similarly, the forward action model is specified as the forward production model, and the forward perceptual model is specified as the forward comprehension model. The comparison of the utterance percept and the predicted utterance percept constitutes self-monitoring.

In Figure 5, the production command constitutes the message that the speaker wishes to convey (see Levelt 1989) and includes information about communicative

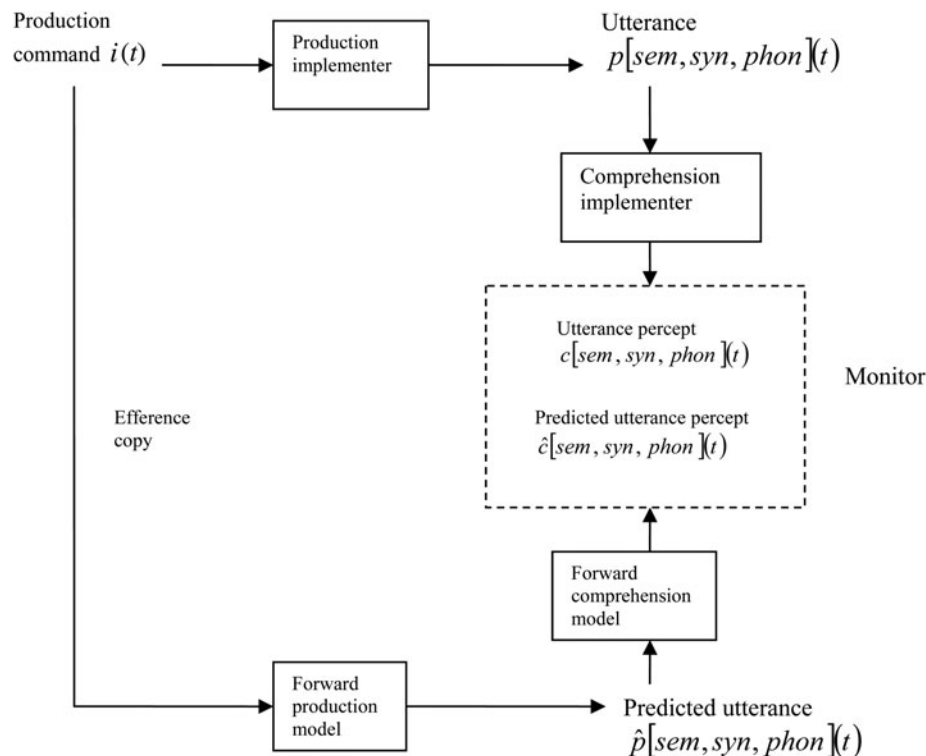


Figure 5. A model of production, using a snapshot of speaking at time  $t$ . The production command  $i(t)$  is used to initiate two processes. First,  $i(t)$  feeds into the production implementer, which outputs an utterance  $p[sem, syn, phon](t)$ , a sequence of sounds that encodes semantics, syntax, and phonology. Notice that  $t$  refers to the time of the production command, not the time at which the representations are computed. In turn, the speaker processes this utterance to create an utterance percept, the perception of a sequence of sounds that encodes semantics, syntax, and phonology. Second, an efference copy of  $i(t)$  feeds into the forward production model, a computational device which outputs a predicted utterance. This feeds into the forward comprehension model, which outputs a predicted utterance percept (i.e., of the predicted semantics, syntax, and phonology). The monitor can then compare the utterance percept and the predicted utterance percept at one or more linguistic levels (and therefore performs self-monitoring).

force (e.g., interrogative), pragmatic context, and a nonlinguistic situation model (e.g., Sanford & Garrod 1981). In addition, Figure 5 does not merely differ from Figure 2 in terminology, but it also assumes structured linguistic representations, such as  $p[sem, syn, phon](t)$  rather than  $a(t)$ . As we have noted, language processing appears to involve a series of intermediate representations between message and articulation. So Figure 5 is a simplification: We assume that speakers construct representations associated with the semantics, syntax, and phonology of the actual utterance, with the semantics being constructed before the syntax, and the syntax before the phonology (in accord with all theories of language production, even if they assume some feedback between representations).<sup>8</sup> We can therefore refer to these individual representations as  $p[sem](t)$ ,  $p[syn](t)$ , and  $p[phon](t)$ . Note that the mappings from  $p[sem](t)$  to  $p[syn](t)$  and  $p[syn](t)$  to  $p[phon](t)$  involve aspects of the production implementer, but Figure 5 places the production implementer before a single representation  $p[sem, syn, phon](t)$  for ease of presentation. Assuming Indefrey and Levelt's (2004) estimates (based on single-word production), semantics (including message preparation) takes about 175 ms, syntax (lemma access) takes about 75 ms, and phonology (including syllabification) takes around 205 ms. Phonetic encoding and articulation takes an additional 145 ms (see Sahin et al. 2009, for slightly longer estimates of syntactic and phonological processing).

Finally, speakers use the comprehension implementer to construct the utterance percept. Again, we assume that this system acts on each production representation individually, so that  $p[sem](t)$  is mapped to  $c[sem](t)$ ,  $p[syn](t)$  to  $c[syn](t)$ , and  $p[phon](t)$  to  $c[phon](t)$ ; therefore, Figure 5 is a simplification in this respect as well. Importantly, the speaker constructs her utterance percept for semantics before syntax before phonology. Unlike Levelt (1989), we therefore assume that the speaker maps between representations associated with production and comprehension at all linguistic levels.

The forward production model constructs  $\hat{p}[sem](t)$ ,  $\hat{p}[syn](t)$ , and  $\hat{p}[phon](t)$ , and the forward comprehension model constructs  $\hat{c}[sem](t)$ ,  $\hat{c}[syn](t)$ , and  $\hat{c}[phon](t)$ . Most important, these representations are typically ready before the representations constructed by the production implementer and the comprehension implementer. The speaker can then use the monitor to compare the predicted utterance percept with the (actual) utterance percept at each level (see Fig. 5) when those actual percepts are ready. Thus, the monitor can compare predicted with actual semantics first, then predicted with actual syntax, then predicted with actual phonology. The production implementer makes occasional errors, and the monitor detects such errors by noting mismatches between outputs of the production implementer and outputs of the forward model. It may then trigger a correction (but does not need to do so). To do this, the monitor must of

course be fairly accurate and use predictions made independently of the production implementer itself.

Let us now consider the content of these predictions and the organization of the forward models in more detail using examples. In doing so, we address the obvious criticism that if the speaker is computing a forward model, why not just use that model in production itself? The answer is that the predictions are not the same as the implemented production representations, but are easier-to-compute “impoverished” representations. They leave out (or simplify) many components of the implemented representations, just as a forward model of predicted hand movements might encode coordinates but not distance between index finger and thumb, or a forward model for navigation might include information about the ship’s position and perhaps fuel level but not its response to the heavy swell.

Similarly, the forward model does not form part of the production command. The production command incorporates a conceptual representation that describes a situation model and communicative force. It cannot represent information such as the first phoneme of the word the speaker is to use, because such information is phonological, not conceptual. In addition, the production command does not involve perceptual representations (what it “feels like” to perform an act), unlike the forward comprehension model.

Additionally, the forward model represents rather than instantiates time. For example, a speaker utters *The boy went outside to fly...* and has decided to produce a word corresponding to a conceptual representation of a kite. At this point, she has predicted that the next word will be a definite determiner with phonology /ðe/, and that its articulation should start in 100 ms. (She does not wait 100 ms to make this prediction.) She may also have predicted some aspects of the following word (*kite*) and that it should start in 300 ms.

But apart from the timing, in what sense is this forward model impoverished? The phonological prediction ( $\hat{p}[\text{phon}](t)$ ) might indicate (for example) the identities of the phonemes (/k/, /a/, /I/, /t/) and their order, but not how they are produced. So when the speaker decides to utter *kite*, she might simply look up the phonemes in a table and associate them with the numbers 1, 2, 3, and 4. Importantly, she does not necessarily have the prediction of /k/ ready before the prediction of /t/. Alternatively, she might look up the first phoneme, in which case the forward model would include information about /k/ only.

Similarly, the syntactic prediction ( $\hat{p}[\text{syn}](t)$ ) might include the grammatical category of noun, but not whether the noun is singular or plural (or its gender, in a gender-marking language). The speaker might simply look up the information that a flyable object is likely to be a noun. This information then suggests that the word should occur at particular positions: for instance, following a determiner. In addition, it is not necessary that the predicted representations are computed sequentially. Although the implemented syntax ( $p[\text{syn}](t)$ ) must be ready before the implemented phonology ( $p[\text{phon}](t)$ ), the syntactic prediction need not be ready before the phonological prediction. For example, the speaker might predict that the kite concept should have the first phoneme /k/ and predict that it should be a noun at the same time, or indeed predict the first phoneme without making any syntactic prediction at all. In summary, we assume that the production system “intervenes” between

the implemented semantics and the implemented syntax, and between the implemented syntax and the implemented phonology, but we do *not* assume intervention in the forward production model.

For example, a speaker might decide to describe a transitive event. At this point, she constructs a forward model of syntax, say  $[NP [V NP]_{VP}]_S$ , where *NP* refers to a noun phrase, *V* a verb, *VP* a verb phrase, and *S* a sentence. This forward model appears appropriate if the speaker knows that transitive events are usually described by transitive constructions, a piece of information assumed in construction grammar (Goldberg 1995), which associates constructions with “general” meanings. The speaker can therefore make this prediction before having decided on other aspects of the semantics of the utterance, thus allowing the syntactic prediction to be ready before the implemented semantics.

At a more abstract level, consider when the speaker wishes to refer to something in common ground (but not highly focused). On the basis of extensive experience, she can predict that the utterance will have the semantics definite nominal, the syntax  $[Det N]_{NP}$  – where *Det* refers to a determiner, *N* a noun, and *NP* a noun phrase – and the phonology starting with /ðe/; she may also predict that she will start uttering the noun in 200 ms.

This approach might underlie choice of syntactic structure during production. For example, speakers of English favor producing short constituents before long ones (e.g., Hawkins 1994). To do this, they might start constructing short and long constituents at the same time but tend to produce short ones first because they are ready first (see Ferreira 1996). However, this appears to be inefficient because it would lead to sharp increases in processing difficulty at specific points (here, when producing the short phrase), and would therefore work against a preference for uniform information density during production (Jaeger 2010, p. 25). It would mean that the long phrase would often be ready much too early, and would incorrectly predict that blend errors should be very common.

Alternatively, the speaker could decide to describe a complex event and a simple event. She uses forward modeling to predict that the complex event will require a heavy phrase and the simple event a light phrase. She then evokes the “short before long” principle, and uses it to convert the simple event into a light phrase (using the production implementer). She can then wait till quite near the end of the phrase before beginning to produce the heavy phrase (again, using the implementer). In this way, she keeps information density fairly constant, prevents blending errors, and reduces memory load.

Just as in action, the speaker “tunes” the forward model based on experience speaking. If she has repeatedly formulated the intention to refer to a kite concept and then uttered the phoneme /k/, she will start to construct an accurate forward model ( $\hat{p}[\text{phon}](t) = /k/$ ) when she next decides to refer to such a concept. If she then constructs an incorrect phonological representation (e.g.,  $p[\text{phon}](t) = /g/$ ), the monitor will likely immediately notice the mismatch between these two representations. If she believes the forward model is accurate, she will detect a speech error, perhaps reformulate, and modify her production implementer for subsequent utterances; if she believes that it may not be accurate, she will not reformulate but will alter her forward model accordingly (cf. Wolpert et al. 2001).

*Evidence from speech production.* There is good evidence for use of forward perceptual models during speech production. In a magnetoencephalography (MEG) study, Heinks-Maldonado et al. (2006) found that the M100 was reduced when people spoke and concurrently listened to their own unaltered speech versus a pitch-shifted distortion of the speech. We assume that they construct a predicted phonological percept,  $\hat{c}[\text{phon}](t)$ . This typically matches their phonological percept ( $c[\text{phon}](t)$ ) and thus suppresses the M100 (i.e., via reafference cancellation). But when the actual speech is distorted, the percept and the predicted percept do not match, and thus the M100 is enhanced. (The M100 could not reflect distorted speech itself as it was not enhanced when distorted speech was replayed to the speakers.) The rapidity of the effect suggests that speakers could not be comprehending what they heard and comparing this to their memory of their planned utterance. Additionally, Tian and Poeppel (2010) had participants produce or imagine producing a syllable, and found the same rapid MEG response in auditory cortex in both conditions. This suggests that speakers construct a forward model incorporating phonological information under conditions when they do not speak (i.e., do not use the production implementer).

Tourville et al. (2008) had participants read aloud monosyllabic words while recording fMRI. On a small proportion of trials, participants' auditory feedback was distorted by shifting the first formant either up or down. Participants compensated by shifting their speech in the opposite direction within 100 ms. Such rapid compensation is a hallmark of feedforward (predictive) monitoring (as correction following feedback would be too slow). Moreover, the fMRI results identified a network of neurons coding mismatches between expected and actual auditory signals. These three studies therefore provide clear evidence for forward models in speech production. In fact, Tourville and Guenther (2011) described a specific implementation of such forward-model-based monitoring in the context of their Directions into Velocities of Articulators (DIVA) and Gradient Order DIVA (GODIVA) models of speech production. However, these data and implementations do not relate to the full set of stages involved in language production.

*Language production and self-monitoring.* In psycholinguistics, well-established accounts of language production (e.g., Bock & Levelt 1994; Dell 1986; Garrett 1980; Hartsuiker & Kolk, 2001; Levelt 1989; Levelt et al. 1999) make no reference to forward modeling, and instead debate the operations of the production implementer (see top line in Fig. 4). They tend to assume that self-monitoring uses the comprehension system. Levelt (1989) proposed that people can monitor what they utter (using an external loop) and thus repair errors. But he noted that they also make many repairs before completing the word, as in *to the ye- to the orange node*, where it is clear that they were going to utter *yellow* (Levelt 1983), and show arousal when they are about to utter a taboo word but do not do so (Motley et al. 1975). Levelt therefore proposed that speakers construct a sound-based representation (originally phonetic, but phonological in Wheeldon & Levelt 1995) and input that representation directly into the comprehension system (using an internal loop). Note that other accounts have assumed monitoring that is more limited (e.g., suggesting that some evidence for monitoring

is in fact due to feedback in the production system; Dell 1986). The accounts do not, however, deny the existence of a comprehension-based monitor.

However, alternative accounts have assumed that at least some monitoring can be “internal” to language production (e.g., Laver 1980; Schlenck et al. 1987; Van Wijk & Kempen 1987; see Postma 2000). Such monitoring could involve the comparison of different aspects of implemented production – for example, if the process is redundantly organized and a problem is noted if the outputs do not match (see Schlenck et al. 1987). Alternatively, it could register a problem if there is high conflict between potential words or phonemes (Nozari et al. 2011). Our account makes the rather different claim that the monitor compares the output of implemented production (the utterance percept) with the output of the forward model (the predicted utterance percept).<sup>9</sup>

Of course, speakers clearly can perform comprehension-based monitoring using the external loop and indeed may be able to perform it using the internal loop as well. But a purely comprehension-based account cannot explain the data from Heinks-Maldonado et al. (2006) and Tourville et al. (2008). In addition, such an account has difficulty explaining the timing of error detection. To correct *to the ye- to the orange node*, the speaker prepares for  $p[\text{phon}](t)$  for *yellow*, converts it into  $c[\text{phon}](t)$ , uses comprehension to construct  $c[\text{sem}](t)$ , judges that  $c[\text{sem}](t)$  is not appropriate (i.e., it is incompatible with  $p[\text{sem}](t)$  or it does not make sense in the context), and manages to stop speaking, *before* she articulates more than *ye-*. Given Indefrey and Levelt's (2004) estimates, the speaker has about 145 ms plus the time to utter *ye-*, which is arguably less than the time it takes to comprehend a word (e.g., Levelt 1989). Speakers might therefore make use of a “buffer” to store intermediate representations and delay phonetic encoding and articulation (e.g., Blackmer & Mitton 1991), but this is unlikely given that they speed up the process of monitoring and repair when speaking faster (see Postma 2000).

Such findings appear incompatible with a purely comprehension-based approach to monitoring.<sup>10</sup> In addition, Nozari et al. (2011) argued that nonspeakers may be able to use the internal loop (as in Wheeldon & Levelt 1995), but that speakers would face the extreme complexity of simultaneously comprehending different parts of an utterance with the internal and the external loops (see also Vigliocco & Hartsuiker 2002). They also noted that there is much evidence for a dissociation between comprehension and self-monitoring in aphasic patients.

Huetting and Hartsuiker (2010) monitored speakers' eye movements while speakers referred to one of four objects in an array. The array contained an object whose name was phonologically related to the name of the target object. In comprehension experiments, people tend to look at such phonological competitors more than unrelated objects (Allopena et al. 1998). Huetting and Hartsuiker found that their speakers also tended to look at competitors after they had produced the target word. This suggests that they monitored their speech using the comprehension system. They did not, however, look at competitors while producing the target word, which suggests that they did not use a comprehension-based monitor of a phonological representation. Huetting and Hartsuiker's findings therefore imply that speakers first monitor using a forward model (as we propose) but can later perform comprehension-based monitoring.

Accounts using an internal loop imply that phonological errors should be detected before semantic errors (assuming that both forms of detection are equally difficult). In contrast, our account claims that speakers construct the predicted semantic, syntactic, and phonological percepts early. Speakers then construct the semantic percept and compare it with the predicted semantic percept; then they construct the syntactic percept and compare it with the predicted syntactic percept; finally, speakers construct the phonological percept and compare it with the predicted phonological percept. Thus, they should detect semantic errors before syntactic errors, and detect syntactic errors before phonological errors.<sup>11,12</sup>

### 3.2. Covert imitation and forward modeling in language comprehension

We now propose an account of prediction during language comprehension that incorporates the account of prediction during language production (see Fig. 5) in the same way that the account of prediction during action perception (see Fig. 3) incorporates the account of prediction during action (see Fig. 2). This account of prediction during language comprehension assumes that people make use of their ability to predict aspects of their own utterances to predict other people's utterances. Of course, language comprehension involves structured linguistic representations (semantics, syntax, and phonology), and different predictions can be made at different levels. Hence prediction is very powerful, because it is often the case that language is highly predictable at one linguistic level at least. An upcoming content word is sometimes predictable. Often, a syntactic category can be predicted when the word itself cannot. On other occasions, the upcoming phoneme is predictable. We propose that comprehenders make whatever linguistic predictions they can.

We assume that people can predict language using the association route and the simulation route. The association route is based on experience in comprehending others' utterances. A comparable mechanism could be used to predict upcoming natural sounds (e.g., of a wave crashing against rocks). The simulation route is based on experience producing utterances. As in action perception, the simplest possibility is that the comprehender works out what he would say under the circumstances more quickly than the producer speaks, using a forward model. But just as with action perception, he needs to be able to represent what the speaker would say, not what he himself would say, and to do this, he needs to take into account the context. We illustrate the model in Figure 6, in which the comprehender A covertly imitates B's unfolding utterance (at time  $t$ ) and uses forward modeling to derive the predicted utterance percept, which can then be compared with A's percept of B's actual utterance (at time  $t + 1$ ). Note that this account differs from Pickering and Garrod (2007), in which the comprehender simply predicts what he would say (and where these representations are not impoverished). Other-monitoring can take place at different linguistic levels, just like self-monitoring.

We now illustrate this account using a situation in which A (a boy) and B (a girl) have been given presents of an airplane and a kite respectively. B utters *I want to go out and fly the*. It is of course highly likely that B will say *kite*, which has  $p[sem, syn, phon]_B(t+1) = [KITE, \text{noun}, /kaɪt/]$ . The

utterance at time  $t$  is the semantics, syntax, and phonology of *I want to go out and fly the*. To predict the situation at time  $t + 1$ , A covertly imitates B's production of *I want to go out and fly the*, and derives the production command that A would use to produce this utterance. A then derives the production command that A would use to produce the word that B would likely say (*kite*) and runs his forward models to derive his predicted utterance percept. If A feels sufficiently certain of what B is likely to say, A can act on this prediction – for example, looking for a kite before B actually says *kite*. In addition, A can compare his prediction of what B will say with what B actually says using the monitor. In this case, A has no access to B's representations during production, and therefore derives the utterance percept from B's actual utterance. This means that A will access B's phonology before B's semantics. In this respect, other-monitoring is different from self-monitoring.

Importantly, A derives the production command of what A assumes B is likely to say (i.e., *kite*), rather than what A himself would be likely to say (i.e., *airplane*). This is the effect of using context together with the inverse model. It is consistent with the finding that comprehenders often pay attention to the speaker's state of knowledge (e.g., Hanna et al. 2003; Metzing & Brennan 2003). However, comprehenders also show some "egocentric biases" (e.g., Keysar et al. 2000), a finding which is expected given that the comprehender's use of context cannot be perfect. Note also that predictions are driven by the forward production model, not by the production system itself. The production system would normally be too slow, given that the speaker should be at least as aware of what she is trying to say as the listener is. Use of the forward model also tends to cause some co-activation of the production system (as is typically the case when forward models are constructed). Such activation is not central to prediction-by-simulation, but can lead to interference between production and comprehension, and serves as the basis for overt imitation (see "Overt responses" in Fig. 6).

Note that Glenberg and Gallese (2012) recently proposed an Action Based Language (ABL) model of acquisition and comprehension that also uses paired inverse and forward models as in MOSAIC. The primary goal of ABL is to account for the content (rather than form) of language understanding, with language comprehension leading to the activation of action-based (embodied) representations. To do this, they specifically draw on evidence from mirror-neuron systems (see sect. 4).

To assess our account, we discuss the evidence that comprehenders make predictions, that they covertly imitate what they hear, and that covert imitation leads to prediction that facilitates comprehension.

**3.2.1. Evidence for prediction.** A great deal of evidence shows that people predict other people's language (see Kutas et al. 2011 and Pickering & Garrod 2007, for reviews). This evidence is compatible with probabilistic models of language comprehension (e.g., Hale 2006; Levy 2008), models of complexity that incorporate prediction (Gibson 1998), and accounts based on simple recurrent networks (Elman 1990; see also Altmann & Mirkovic 2009). But much of the evidence also provides support for aspects of the account in Figure 6.

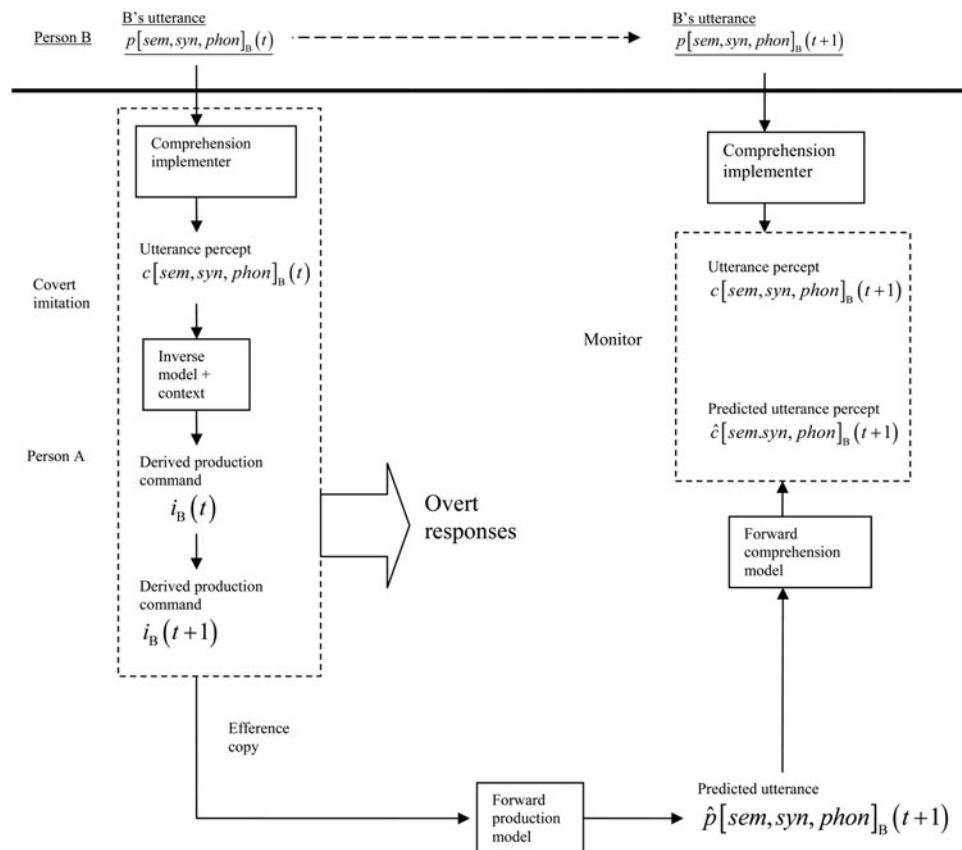


Figure 6. A model of the simulation route to prediction during comprehension in Person A. Everything above the solid line refers to B's unfolding utterance (and is underlined). A predicts B's utterance  $p[sem, syn, phon]_B(t+1)$  (i.e., its upcoming semantics, syntax, and phonology) given B's utterance (i.e., at the present time  $t$ ). To do this, A first covertly imitates B's utterance. This involves deriving a representation of the utterance percept, and then using the inverse model and context (e.g., information about differences between A's speech system and B's speech system) to derive the production command  $i_B(t)$  that A would use if A were to produce B's utterance and from this the production command  $i_B(t+1)$  associated with the next part of B's utterance (e.g., phoneme or word). A now uses the same forward modeling as she does when producing an utterance (see Fig. 4) to produce her predictions of B's utterance and of B's utterance percept (at different linguistic levels). These predictions are typically ready before her comprehension of B's utterance (the utterance percept). She can then compare the utterance percept and the predicted utterance percept at different linguistic levels (and therefore performs other-monitoring). Notice that A can also use the derived production command  $i_B(t)$  to overtly imitate B's utterance and the derived production command  $i_B(t+1)$  to overtly produce the subsequent part of B's utterance (see "Overt responses").

First, prediction occurs at different linguistic levels. Some research shows prediction of phonology (or associated visual or orthographic information). DeLong et al. (2005) recorded ERPs while participants read sentences such as *The day was breezy so the boy went outside to fly...* They showed an N400 effect when the sentence ended with the less predictable *an airplane* than the more predictable *a kite*. The striking finding was that this effect occurred at *a* or *an*. It could not relate to ease of integration but must have involved prediction of the word and its phonological form (i.e., that it began with a consonant). Vissers et al. (2006) found evidence of disruption when a highly predictable word was misspelled, presumably because it clashed with the predicted orthographic representation of the correct word.

Other experiments show prediction of syntax. Van Berkum et al. (2005) found disruption when Dutch readers and listeners encountered an adjective that did not agree in grammatical gender with an upcoming, highly predictable noun. Staub and Clifton (2006) found that people read *or the subway* faster after *The team found either the train ...* than after *The team took the train ...*

In fact, *either* makes the sentence more predictable by ruling out an analysis in which *or* starts a new clause. Similarly, early syntactic anomaly effects in the ERP record are affected by whether the linguistic context predicts a particular syntactic category for the upcoming word or whether the linguistic context is compatible with different syntactic categories (Lau et al. 2006), and reading times are affected by predicted syntactic structure associated with ellipsis (Yoshida et al. 2013).

Clear evidence for semantic prediction comes from eye-tracking studies in which participants listened to sentences while viewing arrays of objects or depictions of events. They started looking at edible objects more than at inedible objects while hearing *the man ate the* (but not when *ate* was replaced with *moved*; Altmann & Kamide 1999). These predictive eye movements do not just depend on the meaning (or lexical associates) of the verb, but are affected by properties of the prior context (Kaiser & Trueswell 2004; Kamide et al. 2003) or other linguistic information such as prosody (Weber et al. 2006). People also predict the upcoming event as well as the upcoming referent (Knoeferle et al. 2005).

Some of these studies do not clearly demonstrate that the predictions are used more rapidly than would be possible with the production implementer. The eye-tracking studies reveal faster predictions, but they may show prediction of semantics (e.g., edible things) rather than a word (e.g., *cake*). However, recent MEG evidence shows sensitivity to syntactic manipulations in little over 100 ms, in visual cortex (Dikker et al. 2009; 2010). For example, the M100 was affected by predictability when the upcoming word looked like a typical noun (e.g., *soda*) but not when it did not (e.g., *infant*). Presumably, these results cannot be due to integration, because activation of the grammatical category of this word (as part of the process of lexical access) could not occur so rapidly or in an area associated with visual form. Instead, the comprehender must predict both syntactic categories and the form most likely associated with those categories, then match those predictions against the upcoming word. Given that syntactic processing does not take place in the visual cortex (or indeed so quickly), these results reflect the visual correlates of syntactic predictions. They suggest that the comprehender constructs a forward model of visual properties (presumably closely linked to phonological properties) on the basis of sentence context and can compare these predicted visual properties with the input within around 100 ms.

Dikker and Pykkänen (2011) found evidence for form prediction on the basis of semantics. Participants saw a picture followed by a noun phrase that matched (or mismatched) the specific item in the picture (e.g., an apple) or the semantic field (e.g., a collection of food). They found an M100 effect in visual cortex associated with matching the specific item but not the semantic field, suggesting that participants predicted the form of the specific word.

Kim and Lai (2012) conducted a similar study to Vissers et al. (2006) and found a P130 effect for contextually supported pseudowords (e.g., ... *bake a ceke*) but not for non-supported pseudowords (e.g., *bake a tont*). In contrast, an N170 effect occurred for non-supported pseudowords (and nonwords). The N170 may relate to lexical access, but the P130 occurs before lexical access can have occurred and again appears to reflect a forward model, in which the comprehender predicts the form of the word (*cake*) and matches the input to that form.<sup>13</sup> In conclusion, these four studies support forward modeling, but they do not discriminate between prediction-by-simulation and prediction-by-association.

**3.2.2. Evidence for covert imitation.** Much evidence suggests that comprehenders activate mechanisms associated with aspects of language production. As we have noted, there appear to be integrated circuits associated with production and comprehension (Pulvermüller & Fadiga 2010). For example, the lateral part of the precentral cortex is active when listening to /p/ and producing /p/, whereas the inferior precentral area is active when listening to /t/ and producing /t/ (Pulvermüller et al. 2006; see also Vigneau et al. 2006; Wilson et al. 2004). We have also noted that tongue and lip muscles are activated during listening to speech but not other sounds (Fadiga et al. 2002; Watkins et al. 2003). More specifically, Yuen et al. (2010) found that listening to incongruent /t/-initial distracters leaves articulatory traces on simultaneous production of /k/ or /s/ phonemes, in the form of increased alveolar

contact. Furthermore, this effect only occurred with incongruent distracters and not with distinct but congruent distracters (e.g., /g/-initial distracters when producing /k/). These results suggest that perceiving speech results in selective, covert, and automatic activation of the speech articulators. Note that these findings show activation of the production implementer (not a forward model).

There is also much evidence for both overt imitation and overt completion. Speakers tend to imitate the speech of other people after they have comprehended it (see Pickering & Garrod 2004), and to repeat each other's choice of words and semantics (Garrod & Anderson 1987), syntax (Branigan et al. 2000), and sound (Pardo 2006). Such imitation can be rapid and apparently automatic; for instance, speakers are almost as quick imitating a phoneme as they are making a simple response to it (Fowler et al. 2003). Speakers also tend to complete others' utterances. For example, Wright and Garrett (1984; see also Peterson et al. 2001) found that participants were faster at naming a word that was syntactically congruent with prior context than a word that was incongruent (even though neither word was semantically appropriate). Moreover, people regularly complete each other's utterances during dialogue (e.g., 1a-c presented in sect. 1.1); see, for example, Clark and Wilkes-Gibbs (1986). Rapid overt imitation and overt completion are of course compatible with prior covert imitation (see "Overt responses" in Fig. 6).

**3.2.3. Evidence that covert imitation facilitates comprehension via prediction.** The previous sections presented evidence that comprehenders make rapid predictions and that they covertly imitate what they hear. But are imitating and predicting causally linked in the way suggested in Figure 6? The evidence for prediction could involve the association route. In addition, covert imitation of language could be used postdictively, to facilitate memory (as a component of rehearsal) or to assist when comprehension leads to incomplete analyses or fails to resolve an ambiguity (see Garrett 2000).

Recent evidence, however, suggests that covert imitation drives predictions that facilitate comprehension. Adank and Devlin (2010) used fMRI to show that during adaptation to time-compressed speech there was increased activation in the left ventral premotor cortex, an area concerned with planning articulation. This suggests that participants covertly imitated the compressed speech as part of the adaptation process that facilitates comprehension. Adank et al. (2010) found that overt imitation of sentences in an unfamiliar accent facilitated comprehension of subsequent sentences in that accent, in the context of noise. This suggests that overt imitation adapts the production system to an unfamiliar accent and therefore that the production system plays an immediate causal role in comprehension.

Ito et al. (2009) manipulated listeners' cheeks as they heard words on a continuum between *had* and *head*. When the skin of the cheek was stretched upward, listeners reported hearing *head* in preference to *had*; when the skin was stretched downward, they reported hearing *had* in preference to *head*. Because production of *had* requires an upward stretch of cheek skin and production of *head* a downward stretch, the results suggest that proprioceptive feedback from the articulators causally affected comprehension (see also Sams et al. 2005). These results could conceivably be postdictive, perhaps relating to reconstruction occurring

during self-report. Clearer evidence comes from Möttönen and Watkins (2009), who used repetitive transcranial magnetic stimulation (rTMS) to temporarily disrupt specific articulator representations during speech perception. Disrupting lip representations in left primary motor cortex impaired categorical perception of speech sounds involving the lips (e.g., /ba/-/da/), but not the perception of sounds involving other articulators (e.g., /ka/-/ga/). Furthermore, D'Ausilio et al. (2009) found that double-pulse TMS administered to the part of the motor cortex controlling lip movements speeded up and increased accuracy of responses to lip-articulated phonemes, whereas TMS administered to the part of the motor cortex controlling tongue movements speeded up and increased accuracy of responses to tongue-articulated phonemes. More recently, D'Ausilio et al. (2011) had participants repeatedly hear a pseudoword (e.g., *birro*) and used TMS to reveal immediate appropriate articulatory activation (associated with *rr*) if they heard the first part of the same word (*bi*, when co-articulated with *rro*) than if they heard the first part of a different word (*bi*, when co-articulated with *ffo*). Thus, covert imitation facilitates speech recognition as it occurs and before it occurs.

A different type of evidence comes from Stephens et al. (2010), who correlated cortical blood-oxygen-level-dependent (BOLD) signal changes between speakers and listeners during the course of a narrative. There was aligned neural activation in many cortical areas at different lags. Sometimes the speaker's neural activity preceded that of the listener, but sometimes the listener's activity preceded that of the speaker. Importantly, listeners whose activity preceded that of the speaker showed better comprehension, suggesting that covert imitation led to prediction and that this prediction facilitated comprehension.

Finally, speakers may use the production system to predict upcoming words (and events) in relation to scenes. In "visual world" experiments, participants activate the phonology associated with the names of the objects (see

Huetting et al. 2011). For example, Huetting and McQueen (2007) had participants listen to a sentence and found that they looked at a picture whose name was phonologically related to a target word (cf. Allopenna et al. 1998) when they viewed the pictures for 2–3 s before hearing the target word but not when they viewed the pictures for 200 ms. In the former case, they presumably had enough time to access the phonological form of the name of the picture.

These studies therefore show that the results of covert imitation have immediate effects on comprehension as a result of prediction. Moreover, we have shown that covert imitation and prediction take place at many linguistic levels. Together, all of these findings provide support for the model of prediction-by-simulation in Figure 6. Of course, comprehenders may also perform prediction-by-association, just as they can for predicting nonlinguistic events.

### 3.3. Interactive language

Interactive conversation is a highly successful form of joint activity. It appears to be very complex, with interlocutors having to switch between production and comprehension, perform both acts at once, and develop their plans on the fly (Garrod & Pickering 2004). Just as we explained joint actions by combining the accounts of action and action perception (see Fig. 4), so we explain conversation by combining the accounts of language production and comprehension (as in Figs. 5 and 6).

Figure 7 shows how both A and B can predict B's upcoming utterance (using prediction-by-simulation). A comprehends B's current utterance and then uses covert imitation and forward modeling; B formulates his forthcoming production command and uses forward modeling based on that command. If A and B are successful, they should make similar predictions about B's upcoming utterance, and they can use those predictions to coordinate (i.e., have a well-organized conversation). Note that they can both compare their predictions with B's forthcoming

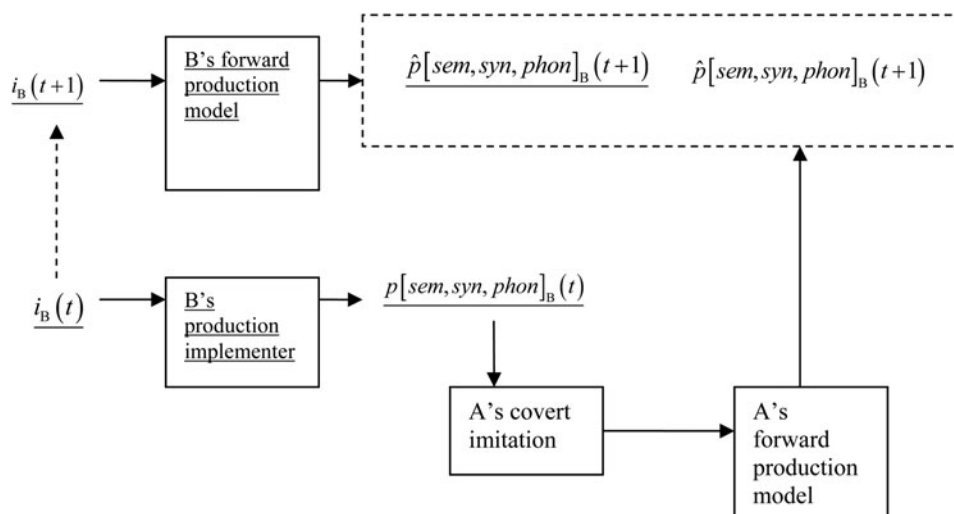


Figure 7. A and B predicting B's forthcoming utterance (with B's processes and representations underlined). B's production command  $i_B(t)$  feeds into B's production implementer and leads to B's utterance  $p[sem, syn, phon]_B(t)$ . A covertly imitates B's utterance and uses A's forward production model to predict B's forthcoming utterance (at time  $t+1$ ). B simultaneously constructs the next production command (the dotted line indicates that this command is causally linked to the previous action command for B but not A) and uses B's forward production model to predict B's forthcoming utterance. If A and B are coordinated, then A's prediction of B's utterance and B's prediction of B's utterance (in the dotted box) should match. Moreover, they may both match B's forthcoming (actual) utterance at time  $t+1$  (not shown).

utterance when produced, with A using other-monitoring and B using self-monitoring. In addition, A and B can also predict A's forthcoming utterance (so both A and B predict both A and B). Of course, these predictions will be related to A's and B's predictions of B's utterance (e.g., both of them might predict both A's upcoming word and B's response following that word), in a way that will reduce the difficulty of making two predictions.

Our account can explain how interlocutors can be so well coordinated – for example, why intervals between turns are so close to 0 ms (Sacks et al. 1974; Wilson & Wilson 2005) and why interlocutors are so good at using the content of utterances to predict when they are likely to end (de Ruiter et al. 2006). Moreover, it accords with the treatment of dialogue as coordinated joint activity, in which partners are able to take different roles as appropriate (Clark 1996). It can also explain the existence and speed of completions, overt imitation (e.g., Branigan et al. 2000; Fowler et al. 2003; Garrod & Anderson 1987), and (assuming links between intentions) rapid complementary responses (as in answers to questions).

We illustrate with the following extract (adapted from Howes et al. 2011):

- 2a – A: ... and then we looked along one deck, we were high up, and down below there were rows of, rows of lifeboats in case, you see,  
 2b – B: –there was an accident  
 2c – A: –of an accident

In (2b–c), B speaks at the same time as A and has a similar understanding to A. B interrupts A, and it is clear that B must be as ready to contribute as A. Because B completes A's utterance without delay, it would not be possible for B to produce (2b) by comprehending (2a) and then preparing a response “from scratch,” as traditional “serial monologue” accounts assume (see Fig. 1). Instead, we assume that B covertly imitates A's utterance, determines A's current production command, determines A's forthcoming production command, and produces an overt completion (see “Overt responses” in Fig. 6). Thus B's response is time-locked to A's contribution. In fact, (2b) is different from A's own continuation (2c). The two continuations are syntactically different (though both grammatical) but semantically equivalent, thereby indicating that prediction can occur differently at different linguistic levels. Note that prediction-by-association might allow B to predict A's continuation, but would not explain the rapidity of B's response, as B would also have to produce the continuation “from scratch.”

During conversation, interlocutors tend to become aligned with each other at different linguistic levels, and such alignment appears to underlie mutual understanding (Pickering & Garrod 2004). Our account can help explain this process, because the close link between production and comprehension leads to tightly yoked representations for comprehension and production, and allows those representations to be used extremely rapidly (see Garrod & Pickering 2009). Note, however, that the relationship also works the other way: Prediction during comprehension is facilitated when the interlocutors are well-aligned, because the comprehender is more likely to predict the speaker accurately (and the speaker is more likely to predict the comprehender's response, as in question-answering). One effect of this is that B's prediction of what A is going to say is more likely to accord with what

B would be likely to say if B spoke at that point. In other words, B's prediction of B's completion becomes a good proxy for B's prediction of A's completion, and so there is less likelihood of an egocentric bias.<sup>14</sup> In fact, linguistic joint action is more likely to be successful and well-coordinated than many other forms of joint action, precisely because the interlocutors communicate with each other and share the goal of mutual understanding.

#### 4. General Discussion

Our accounts of comprehension and dialogue assign a central role to simulation. We discuss three aspects of simulation: the relationship between “online” and “offline” simulation, between prediction-by-simulation and prediction-by-association, and between simulation and embodiment. We conclude by explaining how our account provides an integrated theory of production and comprehension.

We have focused on online simulation, when the comprehender wishes to predict the speaker in real time. However, our notion of simulation is compatible with the simulation theory of mind-reading (Goldman 2006; see Gordon 1986), which is primarily used to explain offline understanding of others. In our account, the comprehender “enters” the simulation during covert imitation, and “exits” after constructing the predicted utterance percept (see Fig. 6). As in our account, Goldman assumed that people covertly imitate as though they were character acting – attempting to resemble their target as much as possible, and then running things forward as well as they can. This means that the derived action command is “supposed” to be the action command of the target, but it incorporates any changes that are required because of bodily differences. (I can walk like Napoleon by putting my hand inside my jacket and seeing how this affects my gait, but I cannot shrink.) In addition, the perceiver may fail to derive the actor's action command correctly, in which case her covert imitation is biased toward her own proclivities.

The important difference between such accounts and ours is that they do not assume forward models and therefore assume that covert imitation uses the action implementer (but inhibiting overt responses). This may be appropriate for offline reasoning but is too slow for prediction (see Goldman 2006, pp. 213–17; Hurley 2008b). Goldman's account uses simulation as an alternative to constructing a theory of the other person's mind. In contrast, our account uses simulation to facilitate processing, which is particularly important when behavior is rapid (as in Grush 2004; Prinz 2006). Clearly, this is the case for language processing.

However, prediction-by-simulation can also be applied offline as part of the process of thinking and planning (as indeed can prediction-by-association). For example, a speaker might think about the likely consequences of producing a particular utterance, both for her own subsequent utterances and perhaps more important for the responses that addressees are likely to produce. She might do this by constructing a predicted utterance percept, using forward modeling. She could also construct an utterance percept (without articulating), using the production implementer and comprehension implementer (see top right of Fig. 5 and discussion in sect. 3.1), as she would typically have enough time to do so. Assuming co-activation, offline predictions may often involve both the production

implementer and forward modeling. See Pezzulo (2011a) for a related discussion.

Our account assigns a central role to prediction-by-simulation, but it assumes that language comprehension and dialogue also involve prediction-by-association. We propose that comprehenders will emphasize simulation when they are (or appear to be) similar to the speaker because simulation will tend to be accurate. These similarities might relate to cultural or educational background or dialect, or, alternatively, to speed or style of language processing. In addition, simulation will be emphasized during dialogue because the interlocutors will tend to become aligned (Pickering & Garrod 2004), and simulation will tend to persist among those in close relationships (who continue to be aligned). In addition, simulation may also be primed during dialogue, because the fact that the comprehender also has to speak may activate mechanisms associated with production. In contrast, prediction-by-association will be emphasized when the comprehender is less similar to the producer, as for example when the comprehender is a native adult speaker of the language and the producer is a nonnative speaker or a child, or when the comprehender does not have the opportunity to speak (as in reading).

We therefore assume that comprehenders emphasize whichever route is likely to be more accurate (given that they should both be fast enough). It may also be that prediction-by-association is more accurate for simple, “one-step” associations between a current and a subsequent state. For example, people can straightforwardly predict that a person who looks confused is likely to respond slowly. In contrast, prediction-by-simulation is likely to be more complex, because it makes use of the structure inherent in the speaker’s own production mechanisms.

Of course, comprehenders may combine prediction-by-simulation and prediction-by-association. They make use of the same representational vocabulary and hence the mental states are the same; the association route simply involves a different (and more straightforward) set of mappings than the simulation route. Informally, for example, if I see that you are about to speak, I can predict your utterances by combining my experiences of how people like you have spoken and my experiences of how I have spoken under similar circumstances.

There is a lot of current interest in the extent to which language is embodied (see Barsalou 1999; Fischer & Zwaan 2008). Such literature focuses on embodiment of content, in which the conceptual content of language is represented in “modal” (i.e., action-based or perceptual) terms (e.g., *kick* is represented in terms of the movements associated with kicking). It is supported by strong evidence from behavioral experiments (e.g., Glenberg & Kaschak 2002) and cognitive neuroscience (e.g., Desai et al. 2010). In contrast, our account is concerned with embodiment of form, which Gallese (2008) called the vehicle level. It assumes that comprehension involves aspects of production, which is a form of action; by definition, production is embodied at the form level. Interestingly, Glenberg and Gallese (2012) used covert imitation and prediction in an account primarily concerned with content embodiment. They explained why representational gesture tends to co-occur with speech by arguing that speaking activates the corresponding action and that the need to perform the action of articulation prevents the inhibition of related gestural actions (see Hostetter & Alibali 2008).

Both our account and embodied accounts seek to abandon the “cognitive sandwich” (Hurley 2008a). Our account assumes that producers use comprehension processes and comprehenders use production processes, whereas embodied accounts assume that producers and comprehenders use perceptual and motor representations associated with the meaning of what they are communicating. Our account does not require such embodiment but is compatible with it.

## 5. Conclusion

Traditional accounts of language assume separate processing “streams” for production and comprehension. They adopt the “cognitive sandwich,” a perspective that is incompatible both with the demands of communication and with extensive data indicating that production and comprehension are tightly interwoven. We therefore propose an account of language processing that abandons the cognitive sandwich. This account assumes a central role to prediction in language production, comprehension, and dialogue. By building on research in action and action perception, we propose that speakers use forward models to predict aspects of their upcoming utterances and listeners covertly imitate speakers and then use forward models based on their own potential utterances to predict what the speakers are likely to say. The account helps explain the rapidity of production and comprehension and the remarkable fluency of dialogue. It thereby provides the basis for a psychological account of human communication.

## ACKNOWLEDGMENTS

We thank Dale Barr, Martin Corley, Chiara Gambi, and Laura Menenti for their comments, and acknowledge support of ESRC Grants RES-062-23-0376 and RES-060-25-0010.

## NOTES

1. The meat is “amodal” in the sense that its representations are couched in terms of abstract symbols rather than in terms of bodily movements (see section 4).

2. Nothing hinges on this particular “traditional” set of levels. For example, it may be correct to distinguish logical form from semantics, or phonetics from phonology.

3. Note that a mapping from semantics to phonology would be a production process, and a mapping from phonology to semantics would be a comprehension process. Some researchers argue that levels can be “skipped” in comprehension (e.g., Ferreira 2003). But mappings between phonology and semantics also occur for other reasons: for example, to express the relationship between emphasis (represented in the message level) and phonological stress, or between meaning and sound in sound symbolism.

4. We assume that prediction is separate from action or perception – that the processes involved in predicting action or perception can at least in principle be distinguished from action or perception itself. In this respect, our account differs from some theories such as that by Elman (1990).

5. Our *forward action model* corresponds to Wolpert’s *forward dynamic model*, and our *forward perception model* corresponds to his *forward output model*.

6. The perceiver also has to accommodate differences in perspective (e.g., when the actor is facing the perceiver). This type of accommodation is less relevant to (spoken) language, so we do not refer to it again.

7. Mirror neurons fire during both action and perceiving an action (Di Pellegrino et al. 1992), and they are of course

compatible with covert imitation during perception. Most evidence for mirror neurons is indirect in humans (e.g., activation of action areas during perception), but Mukamel et al. (2010) used intracranial electrodes to demonstrate widespread mirror activity in Broca's area of an epileptic patient.

8. We assume that speakers implement a level of semantics during production that is distinct from the production command. The production command includes a situation model that incorporates nonlinguistic information, whereas semantics is more akin to an "LF" level of representation (e.g., incorporating quantifier scope).

9. In fact, Wijnen and Kolk (2005) briefly speculated about the possible use of forward and inverse models in monitoring, making reference to Wolpert's proposals.

10. Note that Levelt (1989) assumed that there is appropriateness monitoring that takes place over semantic representations, and that there is no loop based on syntactic representations.

11. The predicted utterance percept must be represented similarly to the utterance percept, in order that they can be compared. Thus, we might expect speakers to have some awareness of the predicted utterance percept as well as the utterance percept. One possibility is that tip-of-the-tongue states constitute awareness of the forward model (in cases when the production implementer fails) rather than incompletely implemented production. For example, the speaker may compute the forward model for the first phoneme (e.g., Brown & McNeill 1966) or grammatical gender (Vigliocco et al. 1997).

12. Some evidence suggests that inner speech may be impoverished (Oppenheim & Dell 2008; 2010; though cf. Corley et al. 2011). An intriguing possibility is that such impoverishment reflects forward modeling rather than an abstract phonological representation constructed by the production implementer.

13. Note that Kim and Lai interpreted their results as involving interaction during early stages of lexical access, but this is not necessary.

14. In fact, our account can explain why completions can be compatible with the perspective of either of the interlocutors. In (1), B said *But have you ...* and A completed with *burned myself?*, A's completion takes A's perspective (myself). However A could have alternatively said *burned yourself?*, thus taking B's perspective (see sect. 1.1).

## Open Peer Commentary

### It ain't what you do (it's the way that you do it)

doi:10.1017/S0140525X12002488

Kenneth John Aitken

Psychology Department, Hillside School, Aberdour, Fife KY3 0RH, Scotland.  
drken.aitken@btinternet.com

**Abstract:** Knowledge of the complexity of human communication comes from three main sources—(i) studies of the linguistics and neuropsychology of dysfunction after brain injury; (ii) studies of the development of social communication in infancy, and its dysfunction in developmental psychopathologies; and (iii) the evolutionary history of human communicative interaction. Together, these suggest the need for a broad, integrated theory of communication of which language forms a small but critical component.

Pickering & Garrod (P&G) are correct in pointing out problems with treating expressive and receptive language as the only separable, distinct, and sufficient components to human communication. Their own discussion of communication as a more elaborated system involving personal action, action perception,

and joint action as inextricably linked and continuously interacting components is a useful extension to this model. It is an advance on the Markov-type linear analyses to which transcribed language lends itself. Their proposal only begins to touch on the complexities inherent in human communication and its evolution.

P&G have previously proposed that humans are "designed" for dialogue, not for monologue (Garrod & Pickering 2009). Their suggested model moves communicative analysis in this direction but provides a rather simplistic approach that they have contrasted, somewhat quixotically, with an even simpler one.

Recent work on the functional neuroanatomy of language suggests at the least that the expressive and receptive phonological, syntactic, and semantic systems, although closely interlinked, can be disambiguated (see, e.g., Ben Shalom & Poeppel 2008; D'Ausilio et al. 2009; Sidtis & Sidtis 2003). Communicative interaction may also involve many nonlinguistic processes, including proprioception (Sams et al. 2005) and interpersonal timing (Richardson et al. 2007), not inherently linked to communicative intent or content.

Some of the richness and complexity in the systems of human communication have been highlighted through neuropsychological analyses of what have been called *disconnection syndromes* (see Catani & ffytche 2005). The approach has illuminated the importance of many nonlinguistic features as in the analysis of "emotional dysprosodias" (Ross 1981; Van Lancker & Breitenstein 2000).

Functional neuroimaging is demonstrating the roles of distributed neural networks in human communication (Vigneau et al. 2006). The utility of this approach is being extended through the development of explanatory models for conditions such as the autistic spectrum disorders (ASDs; Geschwind & Levitt 2007).

Our capacity to understand the complex neural systems in human communication at both the individual and the dyadic level was limited until the development of functional magnetic resonance imaging. Progress has been a function of the development of technologies sufficient to the task rather than through advances in understanding per se. Methods to enable such investigation of the brain in interaction are being actively developed (Schilbach et al. 2012; Schippers et al. 2010).

**The need for a developmental perspective.** The human infant communicates with caregivers in part because of a range of evolutionary adaptations that are successful in engaging with those around them to ensure their survival and in part because of being reared in an environment that has co-evolved to nurture their use of these adaptations. Language is a relatively recent addition to this process that subsequently enables the rapid transmission of knowledge and culture. This rapid transmission of learning to the infant through acculturation by the caregiver is a process seldom observed in other primates (Tomasello 2008).

The human infant is born largely neonotous but with an altricial capacity to engage with caregivers in nonlinguistic forms of reciprocal interaction. Examples are "interactional synchrony" (Condon & Sander 1974); the bidirectional influence seen in the patterning of interaction (Cohn & Tronick 1988); selective preference for maternal voice (e.g., Hepper et al. 1993); imitation of facial expressions (Meltzoff & Moore 1977), and the infant's ability to track the objects of others' eye movements (Beier & Spelke 2012; Navab et al. 2011). Many sensory systems are well developed and involved in communication from tactile (Stack 2007) to the olfactory (Doucet et al. 2009).

It is becoming clear that these processes are neurobiological in origin with individual variations in function arising in part through transgenerational differences in early experience (see Barrett & Fleming 2011).

Timing and prosody are essential features of communicative interaction, and concepts such as vitality contours of interaction and proto-narrative envelopes are helpful in better describing both verbal and nonverbal interaction (see Stern 2010). Prosody in speech is both language and culture dependent, as is neonatal crying (Mampe et al. 2009), presumably through in utero vocal exposure to maternal inflection patterns.

In describing any contingent linguistic system of communication, we need to characterise the dyadic nonverbal mechanisms that are its prerequisite base.

**The evolution of human communication.** The development of more-complex communication in early Homo sapiens is likely to have paralleled the selection pressure for the earlier birth of less-mature infants that enabled more brain growth to occur after birth and reduced the risks of mortality and morbidity to mother and infant through childbirth (see Falk 2004). In consequence, mothers were required to spend more time in caregiving and became more dependent on other adults to support this process. This was also made easier by the division of labour in the production of food, clothing, and shelter, and by the use of fire for warmth and cooking (see Wrangham 2009).

Human communication cannot be simply reduced to a linguistic means for epistemic exchange. Its ontological (Bråten 1998; Gerhardt 2004) and phylogenetic (Arbib 2012; Denton 2005; Panksepp 2004) origins are in the communication of basic affect such as hunger, thirst, discomfort, threat, and affection (Feldman 2007).

The breadth of differing forms of human communication suggests that it is the functional significance of being able to communicate that is critical, and the specific form that this takes is of secondary significance. In terrestrial species, human language may be unique in its complexity and its ability to convey certain types of information, but there is tremendous variation in what can be conveyed. Similar discussions are also found regarding human music (see, e.g., Fitch 2006), but for the same reasons evolutionary arguments remain speculative.

To deconstruct communication into expressive and receptive language and their associated actions is to deal only with a small aspect of this far more complex but fundamental process (Aitken 2008; Aitken & Trevarthen 1997; Trevarthen & Aitken 2001). Greater understanding of the processes involved in deconstruction is likely to require more detailed analysis with, at the least, the triadic modelling of communication's functional components (Fivaz-Depeursinge & Favez 2006; McHale et al. 2008), and the development of a robust methodology for what is being called "second-person neuroscience" (Przyrembel et al. 2012; Schilbach et al. 2012). Incorporating these broader aspects to communication will also require a broader appreciation of human communication's various components including the contributions of the tactile, the olfactory, and the pansensory.

Many different assessment and assimilation approaches will be needed to facilitate such analyses. Some have been developed and can form the basis for a more comprehensive framework for the understanding of human communication (see Anders et al. 2011; Delaherche et al. 2012).

## Evidence for, and predictions from, forward modeling in language production

doi:10.1017/S0140525X1200249X

F.-Xavier Alario and Carlos M. Hamamé

Laboratoire de Psychologie Cognitive, Aix-Marseille Université & CNRS, 13003 Marseille, France.

francois-xavier.alario@univ-amu.fr

carlos-miguel.hamame@univ-amu.fr

<http://www.univ-provence.fr/wlpc/alario>

[http://www.researchgate.net/profile/Carlos\\_Hamame2/](http://www.researchgate.net/profile/Carlos_Hamame2/)

**Abstract:** Pickering & Garrod (P&G) put forward the interesting idea that language production relies on forward modeling operating at multiple processing levels. The evidence currently available to substantiate this idea mostly concerns sensorimotor processes and not more abstract linguistic levels (e.g., syntax, semantics, phonology). The predictions that follow from the claim seem too general, in their current form, to guide specific empirical tests.

A central aspect of Pickering & Garrod's (P&G's) target article is that language production relies on forward modeling processes. These are explicitly described as involving semantic, syntactic, and phonological representations. Here we will attempt to challenge this aspect of their proposal on two grounds: the evidence available for such a claim, and the predictions that may follow from it.

P&G state that there is good evidence for the use of forward models during speech production. This statement must be qualified or clarified. The experimental evidence put forth relies on quite specific language production situations (e.g., vowel articulation or repeated production of very few simple and similar-sounding words). Such linguistic specificity in the stimuli means that the resulting data may not apply to evidence processes other than articulatory motor control. It is dubious that the evidence provides much information about semantic, syntactic, and possibly phonological processes, either in the speech production implementer or in the feedforward model.

Neurophysiological evidence supporting forward modeling in speech production comes from intracranial-EEG studies showing auditory cortex suppression during vocalization (Flinker et al. 2010; Towle et al. 2008). In the visual system, well-identified motor-visual pathways subserve a similar suppression of sensory activity during eye movements (Sommer & Wurtz 2008). Consequently, auditory suppression during speech is attributed to an efference copy which exerts its influence from Broca's area to the auditory cortex via the inferior parietal lobe (Rauschecker & Scott 2009; Tourville & Guenther 2011). Although there is anatomical evidence for such a pathway (Frey et al. 2008), attempts to test the functionality of this connection during speech have proved inconclusive (Flinker et al. 2010; Towle et al. 2008). The largely unclear matter of which motor-auditory pathway could carry such an efference copy is not considered in the target article. More generally, it is difficult to foresee which pathways may underlie the transmission of an efference copy, should linguistic information be involved (semantics, phonology, and syntax).

P&G also refer to previous theoretical work to support their generalization of feed-forward models beyond sensorimotor processes into linguistic levels. They use as an example the previous generalization of a motor control theory (MOSAIC) to a hierarchical version (HMOSAIC) that controls complex sequences of actions. This parallel has not helped to clarify matters. "The HMOSAIC model suggests that there are multiple levels of representation within the sensorimotor system" (Haruno et al. 2003, p. 11). Hence, the HMOSAIC model seems specific to the sensorimotor system, and not necessarily applicable to more abstract levels of representation and processing, which is a core assumption of P&G's proposal.

Scalp-EEG evidence consistent with the hypothesis that language production is monitored by a general-purpose mechanism can be found in Riès et al. (2011). Using a grammatical gender decision task (a proxy for lexical access) and a standard picture naming task, those authors reported postresponse EEG waves very similar to those linked to response monitoring in nonlinguistic tasks (e.g., error-related negativity). In the speech task, the onset of these waves preceded the onset of overt response. Although this timing feature has sometimes been taken as a signature of efference copy (Gehring et al. 1993), it could also reflect the engagement of internal loop monitoring (Riès et al. 2011).

In the absence of strong evidence for some of P&G's claims on forward modeling in language production, it is appropriate to examine the predictions that follow from them, and to gauge how they might guide future empirical tests.

The only explicitly stated prediction regarding forward modeling in language production is worded in broad terms: "[speakers] should detect semantic errors before syntactic errors, and should syntactic errors before phonological errors" (target article, sect. 3.1, para. 23). Although this statement is clear, there are two requirements for testing the relative ordering of error occurrence: that the errors are unambiguously classified, and that a moment of error detection can be defined and measured. These requirements seem difficult to

meet in the absence of (some form of) overt response. Yet the production of such an overt response would complicate the attribution of the detection to forward modeling versus external loop processes. For timing, a common reference time point is required that is available across utterances. A reasonable proxy in experimental setups is stimulus onset, but for narrative speech or dialogue such an event is not easily defined. Testing this prediction is further complicated because “it is not necessary that the predicted representations are computed sequentially [...] the syntactic prediction need not be ready before the phonological prediction” (sect. 3.1, para. 10). This clearly opens the possibility of reordering the sequence in which the parallel outputs of the implemented production and the forward model are compared.

A different tentative prediction can be constructed from P&G’s proposal. The feedforward model is hypothesized to involve “impoverished representations [that] leave out (or simplify) many components of the implemented representations” (sect. 3.1, para. 6). P&G provide various examples of components that “might” (sect. 3.1, para. 9 onwards) be left out. It is not stated whether such opt-out is circumstantial (i.e., whether a component is left out or not depends on the speech act) or systematic (i.e., a component is always omitted from the feedforward model). In the latter case, omitted components would be susceptible only to external error detection and monitoring, whereas included components should be internally detectable and correctable. Checking whether these general statements are amenable to a specific testable hypothesis, contrasting feedforward with inner and overt speech-monitoring performance (e.g., Oppenheim & Dell 2010), would require more space than this commentary can accommodate.

In short, we submit that the evidence presented by P&G for forward models in language production concerns only limited aspects of this behavior, these being primarily sensorimotor processes (i.e., articulatory processes for speech). No currently available evidence calls for a generalization to more abstract levels. On the other hand, the predictions that may follow from this aspect of P&G’s proposal are, in their current form, too general or unconstrained to guide specific empirical tests. These specific points notwithstanding, P&G’s proposal provides a stimulating impetus for combining psychological and neurophysiological evidence more closely.

#### ACKNOWLEDGMENT

Funding from the European Research Council under the European Community’s Seventh Framework Program (FP7/2007–2013 Grant agreement 263575). Institutional support from “Fédération de Recherche 3C” and the “Brain and Language Research Institute,” both at Aix-Marseille Université. We thank Marieke Longcamp for comments and Dashiell Munding for comments and native proofreading

## How do forward models work? And why would you want them?

doi:10.1017/S0140525X12002506

Jeffrey Bowers

School of Experimental Psychology, University of Bristol, BS8 4JS Bristol, United Kingdom.

[j.bowers@bris.ac.uk](mailto:j.bowers@bris.ac.uk)

<http://www.bristol.ac.uk/expsych/people/jeffrey-s-bowers/>

**Abstract:** The project of coordinating perception, comprehension, and motor control is an exciting one, but I found it hard to follow some of Pickering & Garrod’s (P&G’s) arguments as presented. Consequently, my comment is not so much a disagreement with P&G but a query about the logic of forward models: It is not clear how they are supposed to work, nor why they are needed in this (or many other) contexts, and toward that end I present an alternative idea.

According to Pickering & Garrod (P&G), a key feature of forward models is that they are fast; they allow a system to correct itself much more quickly than is possible on the basis of reafferent feedback. I understand how a forward model allows fast feedback when the feedback is based on the output of the forward model itself (as opposed to conditions in which the output of the forward model is compared to real feedback). But to better understand how this approach works, it needs to be made clearer what additional information is included in the forward model that is not in the production or comprehension systems themselves. Otherwise, there is no point. Furthermore, if indeed forward models have extra information, how is this reconciled with the claim that forward models are “impoverished” compared to the production and comprehension implementers.

I also find it unclear how a forward model speeds things up when the output of a forward model and actual feedback (e.g., proprioceptive, visual, auditory, etc.) are compared. The speed with which the forward model can compute seems useless in these conditions, as the model has to wait for the actual feedback before the comparison process can begin. Furthermore, whenever feedback is involved in correcting motor control for future actions (as opposed to the current action), it is not immediately clear why a slow feedback loop is not supposed to work.

At the same time, it is possible to imagine feedback between levels of a production (or a comprehension) system that would be fast enough to correct errors before overt errors are produced (or before comprehension is compromised). Consider the “predictive coding” model introduced by Grossberg (1980). In his Adaptive Resonance Theory (ART) model, a single unit in layer 2 of the network learns to code for a pattern of activation in layer 1. Critically, the learning between the two layers takes place in both bottom-up connections (such that a pattern of activation in layer 1 learns to activate a given unit in layer 2—what Grossberg calls “instar learning”)—and in top-down connections (such that an activated layer 2 unit learns to activate a pattern in layer 1; what Grossberg calls “outstar learning”). These top-down connections in fact support “prediction,” that is, the layer 2 unit learns to activate those layer 1 units that activated it in the past.

The process of identifying an input involves comparing the bottom-up (input) signal with the top-down (predicted) signal: If the two patterns match (in layer 1) the model, the model goes into a state of resonance (with bottom-up and top-down signals reinforcing one another), and this is taken as evidence that the correct node in layer 2 was indeed activated. If not, there is a mistake that needs to be corrected. The identification of a mistake happens quickly, before any learning takes place (in order to solve the stability-plasticity dilemma, otherwise known as “catastrophic interference”). The important point for present purposes is that ART includes fast prediction within a single system, with no need to posit a separate, parallel forward model. (In fact, the ART system does have a separate parallel system that is engaged in cases of bottom-up/top-down mismatch, but this system does not carry any information about a prediction—it just turns off the layer 2 unit and tells the network to “try again.”)

Could something similar work in the case of speech production? Perhaps a semantic input could activate a lemma unit, and the lemma could feed back to the semantic system, and the model could be confident that the correct lemma was selected if its top-down and bottom-up signals matched. Similar top-down/bottom-up interactions across levels in the speech production system could lead to quick corrections at each stage. And, ultimately, feedback from the actual output (a spoken word in the case of speech production) could play a role in correcting errors (after the fact). Given that predictions are made between levels of the speech production system (and possibly between production and comprehension systems), corrections could presumably be made quickly, and at different stages of the process.

Of course, this sketch of an outline of an idea does not even attempt to address the complexities that are addressed by P&G, and any proposal without a forward model may well be

inadequate. But I'm struggling to see why separate fast forward models are needed (as opposed to feedback between levels within and between production and comprehension systems), and to understand how forward models are thought to be fast whenever they rely on actual feedback.

## Prediction in processing is a by-product of language learning

doi:10.1017/S0140525X12002518

Franklin Chang,<sup>a</sup> Evan Kidd,<sup>b</sup> and Caroline F. Rowland<sup>a</sup>

<sup>a</sup>University of Liverpool, Institute of Psychology, Health and Society, Liverpool L69 7ZA, United Kingdom. <sup>b</sup>The Australian National University, Research School of Psychology, The Australian National University, Canberra 0200, Australia.

Franklin.Chang@liverpool.ac.uk    evan.kidd@anu.edu.au  
crowland@liverpool.ac.uk

<http://www.liv.ac.uk/psychology-health-and-society/staff/franklin-chang/>  
[http://psychology.anu.edu.au/\\_people/people\\_details.asp?recid=594](http://psychology.anu.edu.au/_people/people_details.asp?recid=594)  
<http://www.liv.ac.uk/psychology-health-and-society/staff/caroline-rowland/>

**Abstract:** Both children and adults predict the content of upcoming language, suggesting that prediction is useful for learning as well as processing. We present an alternative model which can explain prediction behaviour as a by-product of language learning. We suggest that a consideration of language acquisition places important constraints on Pickering & Garrod's (P&G's) theory.

Pickering & Garrod (P&G) have done the field a huge favour by conceptualising language processing in terms of an integrated action-perception system, which highlights the centrality of prediction. It seems to us that taking a similar approach to the field of language acquisition would be equally productive. After all, the task of learning a language requires a close integration of oral motor action and sound perception. P&G's model adopts a theory of acquisition from motor learning, in which forward models learn to bridge between action and perception. In this commentary, we explore the implications of P&G's forward model for our understanding of psycholinguistic processes, with a particular focus on language acquisition.

There is now substantial evidence that children use prediction in comprehension in much the same way as adults. For example, Lew-Williams and Fernald (2007) reported that Spanish-learning 3-year-old children can use the grammatical gender of the article (*la/el*) to predict the referent of the next word (*la pelota* vs. *el zapato*) in a looking-while-listening task. Mani and Huettig (2012) have shown that 2-year-olds can use a verb's semantic affordances in a similar manner to adults (e.g., *eat* predicts *cake*), and Borovsky et al. (2012) have reported similar findings in older children and adults. Importantly, the acquisition studies have demonstrated that children's predictive abilities correlate with their knowledge of language. In all of these studies, children (and adults) with bigger productive (and/or receptive) vocabularies tend to be faster and more accurate at prediction during online sentence comprehension. Faster prediction correlates with positive language outcomes longitudinally: Marchman and Fernald (2008) reported that the speed at which 25-month-olds process lexical information correlates with linguistic knowledge 6 years later. These findings suggest that prediction is not only part of the language processing system, but is tightly linked to language acquisition mechanisms.

Scrutinizing the mechanism underlying prediction therefore appears an important priority. Here we compare P&G's model to an alternative language production model, Chang's (2009) Dual-path model, concentrating on how each model explains prediction. By way of example, we consider verb-object affordances

(e.g., *eat-cake*), where both children and adults have been found to predict a verb's object before the object is encountered in speech (Altmann & Kamide 1999). These results are particularly interesting, because Mani and Huettig (2012) found a significant relationship between expressive vocabulary and prediction behaviour in children, which supports both P&G's and the Dual-path model's hypothesis that production representations support prediction.

P&G's model places prediction within a forward model that implements constraints on word order (prediction-as-processing). The model explains verb-object affordances in the following manner: a production command EAT(CAKE) can be derived from hearing the word *eat* and seeing the cake in the visual scene. The command passes through the production forward system to activate the triplet  $p[\text{cake, NP, /cake/}]$ . Crucially, the model predicts that only syntactically and semantically appropriate predictions will be generated. In the *eat-cake* example, the prediction is correct, but is not borne out in every instance. For instance, Kamide et al. (2003) reported that participants made predictive looks to a cabbage after the processing the fragment *The hare will be eaten ...*, where the passivized verb should have restricted looks to an animate agent (e.g., a fox).

In contrast to P&G's model, the Dual-path model acquires syntactic representations within a network that contains separate meaning and sequencing pathways (Elman 1990). Its learning algorithm compares the predicted next word with the actual comprehended next word, and the mismatch or error is used to adjust the model's internal representations (error-based learning, Rumelhart et al. 1986). The Dual-path model is able to explain the phenomena of structural priming as error-based learning (Chang et al. 2006), and this ability requires that prediction-for-learning is constantly taking place during language comprehension. For the *eat-cake* prediction, the input word *eat* activates the concept CAKE because the model's meaning pathway learns associations between words in utterances and elements in messages. Critically, the same word-concept mechanism learns to associate *eaten* with cabbage. This associative word-concept prediction mechanism is different from the structure-sensitive prediction mechanism in the sequencing system, which can explain why *eaten* increases predictive looks to likely agents like the fox in Kamide et al. (2003). Thus, the pathways in the model can explain the different types of prediction. Crucially, the similarity in prediction in children and adults can be explained by the idea that humans are constantly doing prediction-for-learning to adjust their language representations to the input (Kidd 2012; Rowland et al. 2012).

Learning processes can explain prediction in processing, but language acquisition constraints are also critical for learning the syntactic and semantic representations that support prediction in P&G's model. P&G's theory is based on Wolpert et al.'s (2011) theory of motor planning and perception, which uses error-based learning for motor and forward model learning (Jordan & Rumelhart 1992; Plaut & Kello 1999). According to Wolpert et al.'s theory, the forward model is learned by mapping from muscle commands to the perception of one's arm in three-dimensional space. These algorithms work because humans can directly perceive their arm's position. In P&G's theory, the forward model maps from a message-like production command to a triplet including syntax and semantics. This is problematic: We cannot directly perceive syntax and semantics, and hence the learning mechanism in the motor theory cannot explain how P&G's forward model learns to make these language predictions. When error-based learning is used to map from production commands to sentences, Chang (2002) demonstrated that abstract syntax was not always learned unless the model had language acquisition constraints like those in the Dual-path architecture. Therefore, P&G's forward model may need a similar architecture to yield the appropriate predictions.

Language processing theories like P&G's account treat language learning as a peripheral process. We argue that

prediction in processing is actually a by-product of learning. Prediction is a critical component of error-based learning, which is one of the most successful accounts of both motor and language learning.

## Forward modelling requires intention recognition and non-impooverished predictions

doi:10.1017/S0140525X1200252X

Jan P. de Ruiter and Chris Cummins

Department of Psycholinguistics, Bielefeld University, 33501 Bielefeld, Germany.

jan.deruiter@uni-bielefeld.de chris.cummins@uni-bielefeld.de

<http://www.uni-bielefeld.de/lili/personen/jruiter/>

**Abstract:** We encourage Pickering & Garrod (P&G) to implement this promising theory in a computational model. The proposed theory crucially relies on having an efficient and reliable mechanism for early intention recognition. Furthermore, the generation of impoverished predictions is incompatible with a number of key phenomena that motivated P&G's theory. Explaining these phenomena requires fully specified perceptual predictions in both comprehension and production.

We heartily congratulate Pickering & Garrod (P&G) on their outline of a new cognitive architecture that integrates language production and comprehension. What is particularly impressive in their sketch is that it is a comprehensive approach that addresses entire classes of interrelated psycholinguistic phenomena (as opposed to a selected subset of empirical findings) and that it provides natural explanations for especially the time-critical phenomena which have been difficult to explain plausibly and elegantly with our "standard models" (e.g., Dell 1986; Levelt 1989).

Therefore, it is, in our view, all the more urgent that this sketch, or at least its central parts, be further developed into a real computational implementation, one that actually generates the complex behavior that we can now only simulate in our minds on the basis of verbal accounts. Only with such an implementation will we be able to assess the adequacy and accuracy of the provided account, and be able to generate nontrivial and testable predictions that can subsequently be tested using the large arsenal of behavioural and neurocognitive methods now available. We are aware that implementing models is not an easy task, and that implementing this particular architecture will prove to be a challenging exercise. But it is possible to use a piecemeal approach: An obvious simplification is to first develop the model on a miniature language, in a restricted context, using simulated time. Also, it is possible and probably advantageous to employ division of labour by delegating parts of the implementation to different research groups that have complementary expertise.

There are two central aspects of the proposed theory that we would like to comment on, and suggest improvements to.

The first concerns the role of intentions. As P&G note, when predicting the utterance of an interlocutor (i.e., in comprehension), it is essential to have an (early) estimate of the underlying intention. In the HMOSAIC model, this is done by running parallel inverse models. But in modelling verbal interaction, one of the most intractable problems is the complex, seemingly arbitrary, many-to-many mapping of utterances and intentions (see, e.g., Levinson 1983; 1995). We suspect, therefore, that in a model of language production and comprehension (i.e., dialogue processing) this problem is much harder than in the recognition of intentions underlying functional motor behaviour. There are computational models that use Bayesian machine learning procedures to capture the utterance-intention mapping from multimodal interaction corpora (see, e.g., DeVault et al. 2011), but this approach involves computationally expensive and time-consuming offline learning procedures, and the resulting models

are limited to the domain they have been trained on (for an alternative Bayesian approach to attacking this problem that does not involve offline training procedures, see De Ruiter & Cummins 2012). We would urge P&G to prioritize this aspect, as we believe that the success of the proposed approach will be to a large degree dependent on its ability to model intention recognition in dialogue.

The second comment we have involves the use of "impoverished" representations for the efferent copies, and especially the nature of this impoverishment. In P&G's exposition of the theory, the stated reason that the system does not simply use the efferent copies as motor programs is their impoverished nature (target article, sect. 3.1, para. 6). However, as the efferent copy represents the *perceptual* consequences of the motor program (and not the motor program itself), not using them directly as motor programs, in our view, does not need to be motivated at all. It is simply a different type of representation, not suited as a motor program.

A potentially more serious problem with the proposed impoverished nature of the efferent copies is that they do not adequately explain the phenomena they are supposed to. This holds for both comprehension and production. In production, for instance, the cited findings by Heinks-Maldonado et al. (2006), and especially those by Tourville et al. (2008), can only be explained if the efferent copy is fully specified, not merely phonologically but also phonetically. If a speaker knows only which phonemes he is going to produce in what order but not how (in terms of phonetic detail), then the proposed theory would predict that changing the first formant in the auditory feedback (as Tourville et al. did) would have no effect at all.

In language comprehension, the proposed theory assumes that listeners predict what their interlocutor is going to say. Indeed, this appears to be essential for explaining the phenomenon of *close shadowing* (Marslen-Wilson 1973), with delays as short as 250 ms. Also, predictions of utterance content probably underlie the listener's highly accurate anticipation of the end of the speaker's turn as found, for instance, by De Ruiter et al. (2006) and Stivers et al. (2009). But here too, the accuracy obtained from having access to an early but impoverished prediction would not be able to explain the levels of accuracy observed in end-of-turn anticipation in experiments and natural data. Magyari and De Ruiter (2012) found evidence that people are able to predict *when* a turn ends by predicting *how* it ends – that is, with which specific words the turn will end. This suggests that the forward model cannot be lexically impoverished, as suggested by P&G in section 3.1 (para. 9).

This is why we would strongly urge P&G to adopt the assumption that the representations of the predictions, both in production and comprehension are fully specified (perceptual) representations, as Pickering and Garrod (2007) suggested for comprehension.

Finally, we again want to express our support for the exciting approach that P&G have taken with their highly original and thought-provoking outline, and look forward to discussing these issues further.

## Cascading and feedback in interactive models of production: A reflection of forward modeling?

doi:10.1017/S0140525X12002531

Gary S. Dell

Beckman Institute, University of Illinois, Urbana-Champaign, Urbana IL 61820.  
gdell@illinois.edu

**Abstract:** Interactive theories of lexical retrieval in language production assume that activation cascades from earlier to later processing levels,

and feeds back in the reverse direction. This commentary invites Pickering & Garrod (P&G) to consider whether cascading and feedback can be seen as a form of forwarding modeling within a hierarchical production system.

Over the past 20 years, one of the most contentious issues in language production has concerned the degree to which lexical access occurs in discrete stages. Theorists agree that words are retrieved first as semantic-syntactic entities, and then later spelled out in terms of their phonological forms (e.g., Garrett 1975; Kempen & Huijbers 1983). But is the first of these steps – lemma or word-access – entirely separate from the second one – phonological access? Three possibilities are debated. The *discrete-stage* or modular view (Levelt et al. 1999) holds that the first step must be completed before any activation of phonological forms takes place. During the first step of retrieval of the word “cat,” the lemmas for CAT and DOG may both be active, but the phonological forms of these will not be. Not until the first step has completed its selection of CAT can /k/, /æ/, and /t/ gain activation. The *cascade* hypothesis blurs the distinction between the steps by allowing for activation of phonological forms of potential lemmas (e.g., the forms of both “cat” and “dog”) before the first step has been completed. The *interactive* hypothesis permits cascading, but also bottom-up feedback (e.g., Dell 1986). Activated phonological units send activation upwards to lexical units. This loop of cascading and feedback between units at adjacent levels of the system is assumed to operate regardless of whether the lexical access process is engaged in word or phonological access. Currently, there is a considerable amount of evidence for cascading (e.g., Cutting & Ferreira 1999), but little consensus on the degree to which the system is interactive (see Harley 2008, for review).

I suggest that Pickering & Garrod’s (P&G’s) proposed use of forward modeling and the predicted perceptions that result from it map onto the notions of cascading and feedback in interactive models of production. Cascading consists of a prediction by processing level *i* of what needs to be active on the next lower level *i+1*, and feedback from that level delivers the anticipated “sensory” consequences of that prediction back to level *i*.

In P&G’s view, the advance prediction of representational components of an utterance and their sensory consequences allow for each production decision to be coordinated with other decisions. They illustrate by showing how heavy noun phrase (NP) shift and phonological error monitoring could result from this system. Generally speaking, forward modeling during production helps make the many parts of an utterance mesh for accurate fluent speech. That is also the function of cascading and feedback in hierarchical production models, except that representational levels rather than utterance parts are what are being meshed. Cascading of activation to lower levels prepares the way for the construction of representations at those levels. The resulting feedback allows for decisions at the higher representational level to be sensitive to information at the lower level. For example, feedback from phonological forms to the word/lemma level allows for word selection at the higher level to reflect the retrievability of the form (Dell et al. 1997). A phonological form that is more easily available will feed back more activation to its lemma than a form that is difficult to retrieve will. Hence, the system will be biased to select lemmas whose forms will be available. Feedback also enables representations at a lower level to mesh with higher-level information, as seen when feedback from phonological to lexical levels biases the phonological level activations toward lexical outcomes, functioning as a lexical editor (e.g., Nozari & Dell 2009).

P&G emphasize that forward predictions are not actual production representations. They are “easier-to-compute ‘impoverished’ representations” (target article, sect. 3.1, para. 6). This is also true of cascading in interactive models. Units that are active through cascading are less active than they would be if they were committed parts of a representation. Furthermore, unlike committed representational elements, they have yet to be

bound to structural frames, at least in activation-based models that use such frames (e.g., Dell 1986). So, although units activated through cascading may soon be fully part of an utterance’s representation at a particular level, they are not there yet.

P&G have outlined a compelling integrated theory of production and comprehension, showing how each contributes to the other. With this commentary, I invite them to consider whether the notions of cascading and feedback in production are part of the picture, and particularly whether they can be considered to be reflections of the forward modeling system operating between processing levels.

#### ACKNOWLEDGMENT

Preparation of this commentary was supported by NIH DC-000191

## The neurobiology of receptive-expressive language interdependence

doi:10.1017/S0140525X12002543

Anthony Steven Dick<sup>a</sup> and Michael Andric<sup>b</sup>

<sup>a</sup>Department of Psychology, Florida International University, Miami, FL 33199;

<sup>b</sup>Center for Mind/Brain Sciences (CIMEC), The University of Trento, Trento 38122, Italy.

adick@fiu.edu michael.andric@unitn.it  
www.fiu.edu/~adick michaelandric.tumblr.com

**Abstract:** With a focus on receptive language, we examine the neurobiological evidence for the interdependence of receptive and expressive language processes. While we agree that there is compelling evidence for such interdependence, we suggest that Pickering & Garrod’s (P&G’s) account would be enhanced by considering more-specific situations in which their model does, and does not, apply.

The classical Lichtheim–Broca–Wernicke neurobiological model of language proposed distinct neuroanatomical pathways for language comprehension and production. Recent evidence suggests abandoning this model’s classical form, and although there is not yet an established replacement (Dick & Tremblay 2012; Price 2010; 2012 for review), we think much of the data support P&G’s proposal. However, we also think P&G could be clearer about whether there are situations in which their model does not apply. For example, they state that “comprehenders make whatever linguistic predictions they can” (target article, sect. 3.2, para. 1), but this is so broad as to be unfalsifiable.

Neurobiological evidence suggests production and perception system interdependence occurs in specific situations. By highlighting emerging models and findings in the neurobiology of receptive language, we suggest that P&G’s proposal could be fine-tuned to make more-specific, testable predictions.

**Neurobiological evidence for the interdependence of receptive-expressive language in speech perception.** The most widely adopted model of language neurobiology is a dual-stream model analogous to the visual system (Ungerleider & Haxby 1994). Within this model, during receptive language, auditory speech sounds map to articulatory (motor) representations in a dorsal stream and to meaning in a ventral stream (Hickok 2009b; Hickok & Poeppel 2000; 2004; 2007; Rauschecker 2011; Rauschecker & Scott 2009; Rauschecker & Tian 2000; Rogalsky & Hickok 2011). If this is correct, models like P&G’s must account for the way these processing streams interact with the motor system involved in language production.

This problem is easier to solve within the dorsal stream, as many of the same brain regions are active during speech planning and execution, and during speech perception (Callan et al. 2004; Eickhoff et al. 2009; Hickok & Poeppel 2007; Pulvermüller et al. 2006; Vigneau et al. 2006; Wilson et al. 2004). In fact, a primary contention is not *whether* the motor system is recruited

during speech perception but *in what situations* it occurs. Some argue the motor system is essential (D'Ausilio et al. 2009; Iacoboni 2008; Meister et al. 2007), whereas others argue that it is only involved when auditory-only speech is difficult to parse (e.g., during noisy situations, or when discriminating between similar phonemic units; Hickok 2009a; Hickok et al. 2011; Sato et al. 2009; Tremblay & Small 2011).

The latter situation appears to be the case for audiovisual speech perception, when visual information from the lips and mouth is present. Moreover, a forward-modeling architecture consistent with P&G's proposal has been suggested to explain the neurobiology of audiovisual speech perception (Callan et al. 2004; Skipper et al. 2005; Skipper et al. 2007b; van Wassenhove et al. 2005; Wilson & Iacoboni 2006). Here, visual information, temporally preceding the auditory signal by several hundred milliseconds (Chandrasekaran et al. 2009), provides a "forward model" of the speech sound. These models draw on the listener's articulatory representations to provide possible phonetic targets of the talker's speech (Callan et al. 2004; Skipper et al. 2007b; van Wassenhove et al. 2005). Findings that visual speech influences the auditory neural response's latency and amplitude (van Wassenhove et al. 2005), and recruits motor-speech regions (Callan et al. 2004; Dick et al. 2010; Hasson et al. 2007; Sato et al. 2010; Skipper et al. 2005; 2007b; Watkins et al. 2003), support predictive coding via forward models of the kind P&G propose.

#### **Neurobiological evidence for the interdependence of receptive-expressive language in language and gesture comprehension.**

Although the neurobiological evidence for receptive-expressive language interdependence is compelling in speech perception, it is mixed for higher-level language comprehension, which involves brain regions along a ventral language pathway (Binder et al. 2009; Hickok & Poeppel 2007; Vigneau et al. 2006). There is evidence – for example, in processing verbs – that the motor system contributes to understanding, and this is cited to support "motor simulation" theories (Cappa & Pulvermüller 2012; Fischer & Zwaan 2008; Glenberg 2011; Glenberg & Gallese 2012). Notably, some authors interpret these findings without adhering to motor simulation theories (Bedny & Caramazza 2011; Mahon & Caramazza, 2009). Indeed, motor (production) system contribution to language comprehension is a contentious issue (e.g., this was a topic of an organized debate at the 2011 Neurobiology of Language Conference).

Additional evidence suggests that involvement of the motor system is specific to the task. For example, Tremblay et al. (2012) applied repetitive transcranial magnetic stimulation (rTMS) to the ventral premotor cortex during a sentence comprehension task. The rTMS interfered with sentences describing manual actions, but not with other types of sentences, suggesting that predictive motor encoding is not always called upon. Another example is gesture comprehension. Some studies have shown that the act of viewing gestures recruits areas associated with a putative "mirror neuron" system thought to covertly simulate others' actions (Green et al. 2009; Holle et al. 2008; Skipper et al. 2007a; 2009; Willems et al. 2007; Xu et al. 2009), but others show no evidence that this correlates with comprehension (Andric & Small 2012; Dick et al. 2009; 2012; Straube et al. 2011; Willems et al. 2009).

In closing, we note that within P&G's model it may not be necessary to elicit motor activation. For example, P&G state that "embodied accounts assume that producers and comprehenders use perceptual and motor representations associated with the meaning of what they are communicating. Our account does not require such embodiment but is compatible with it" (sect. 4, para. 9). Hence, the model seems able to account for motor activity, or lack of it, during receptive language. If this is the case, P&G should clarify what neurobiological findings could help decide between competing accounts that call upon interdependent receptive and expressive language systems.

## **Intermediate representations exclude embodiment**

doi:10.1017/S0140525X12002555

Guy Dove

Department of Philosophy, University of Louisville, Louisville, KY 40292.

[guy.dove@louisville.edu](mailto:guy.dove@louisville.edu)

<http://louisville.edu/faculty/godove01/>

**Abstract:** Given that Pickering & Garrod's (P&G's) account integrates language production and comprehension, it is reasonable to ask whether it is compatible with embodied cognition. I argue that its dependence on rich intermediate representations of linguistic structure excludes embodiment. Two options are available to supporters of embodied cognition: They can adopt a more liberal notion of embodiment or they can attempt to replace these intermediate representations with robustly embodied ones. Both of these options face challenges.

Pickering & Garrod (P&G) maintain that their integrated approach to language production and comprehension is compatible with, but does not require, embodiment. I argue that it is incompatible. The fundamental role played by intermediate representations that capture phonological, syntactic, and semantic structure rules out embodiment and preserves important aspects of classical cognitive science.

**Integration and embodiment.** There is a clear affinity between P&G's project and that of embodied cognition. The basic idea behind embodied cognition is that cognitive processes are partly constituted by wider bodily structures and processes. Glenberg (2010, p. 586) characterized it as the claim that "all psychological processes are influenced by body morphology, sensory systems, motor systems, and emotions." Such influence clearly requires an intimate relationship between action, perception, and cognition.

Traditionally, researchers have assumed that language processing is inherently modular; they have been committed to what Hurley (2008a) calls the *classical sandwich*. On this view, central cognition (the meat) intercedes between action and perception (the slices of bread). P&G argue that language production and comprehension are forms of action and action perception respectively. As they see it, receivers of linguistic messages actively compute action representations during perception to help them predict what they are about to perceive. Similarly, producers of a linguistic message actively compute perception representations to help them predict sensory feedback from their ongoing action. In violation of the classical sandwich, comprehension often involves production processes and production often involves comprehension processes.

One of P&G's primary theoretical innovations is their use of forward and inverse modeling to account for the dynamic nature of language processing. This clearly fits with the central role played by perceptual and motor simulation in many accounts of embodied cognition (for reviews, see Barsalou 2008; Kemmerer 2010; Martin & Zwaan 2008). It also fits with the more recent suggestion that prediction is important to guiding action and perception (Gallese 2009).

**The problem posed by intermediate representations.** A core aspect of P&G's theory does not fit with embodied cognition: its reliance on disembodied representations. The problem begins with their acknowledgement that language is special: "Unlike many other forms of action and perception, language processing is clearly structured, incorporating well-defined levels of linguistic representation such as semantics, syntax, and phonology" (target article, sect. 1.3, para. 9). To handle the linguistic structure at these three levels, they posit "a series of intermediate representations between message and articulation" (sect. 3.1, para. 3). These intermediate representations are central to their account. Indeed, P&G define production processes as those that map "higher" linguistic representations to "lower" ones and comprehension processes as those that map "lower" linguistic representations to "higher" ones.

One of the difficulties facing any attempt to assess embodied cognition is that it has been associated with several distinct theses (Anderson 2003; Shapiro 2011; Wilson 2002). There is, however, good reason to think that P&G's intermediate representations are incompatible with most versions of embodiment. Obviously, any appeal to representations excludes radical anti-representational forms of embodied cognitive science (Chemero 2009). Less radical forms of embodied cognitive science generally assume that embodiment requires, at a minimum, grounding in modality-specific input/output systems. Pezzulo et al. (2011, p. 3) outline a core feature of this grounding: "Perhaps the first and foremost attribute of a grounded computational model is the implementation of cognitive processes... as depending on *modal* representations and associated mechanisms for their processing... rather than on amodal representations, transductions, and abstract rule systems." P&G's intermediate representations clearly fail to meet this criterion.

Traditional cognitive science posits amodal representations for a reason: They provide a means of integrating information associated with distinct perceptual and motor modalities. Psycholinguists often argue that amodal representations are needed for language processing because linguistic structure transcends the particulars of the various modalities associated with production and comprehension (e.g., Jackendoff 2002; 2007; Pinker 2007). On the syntactic front, the well-known structural similarity of signed and spoken languages is taken to provide further support for this claim (Goldin-Meadow 2005; Poizner et al. 1987).

Although the need for amodal representations has typically been formulated against the assumption that production and comprehension are separate processes, this background assumption is not necessary. Indeed, as P&G show, amodal representations can serve as a bridge for the ongoing interaction between production and comprehension. To mangle a cliché, P&G provide a way to avoid throwing the meat out with the sandwich by identifying an important role for the sort of amodal representations posited by traditional cognitive science within a non-modular account of language processing.

**Conclusion.** P&G's appeal to intermediate representations leaves supporters of embodied cognition with something of a dilemma. On the one hand, they could try to liberalize the notion of embodiment in order to encompass such representations. This move is not without precedent. Meteyard et al. (2012), for example, argue that researchers need to consider the possibility that cognition is *weakly embodied* because it involves supramodal representations that capture associations between distinct sensorimotor systems (Barsalou et al. 2003; Damasio & Damasio 1994; Gallese & Lakoff 2005). The obvious danger of this strategy is that it could erode the force and novelty of the thesis that cognition is embodied. On the other hand, they could try to offer more robustly embodied accounts of phonological, syntactic, and semantic knowledge. This strategy faces a general challenge: In order to eliminate the need for amodal representations at a given level, it is not enough to show that some phenomena at that level can be handled in an embodied fashion. Instead, what needs to be shown is that *all* of the phenomena at that level can be handled in this way (Toni et al. 2008). This sets the bar very high, and we can reasonably doubt that it is achievable. As matters stand, neither of these options seems particularly promising.

## The role of action in verbal communication and shared reality

doi:10.1017/S0140525X12002567

Gerald Echterhoff

Social Psychology Group, Department of Psychology, University of Münster, D-48149 Münster, Germany.

[g.echterhoff@uni-muenster.de](mailto:g.echterhoff@uni-muenster.de)

<http://geraldlechterhoff.com>

**Abstract:** In examining the utility of the action view advanced in the Pickering & Garrod (P&G) target article, I first consider its contribution to the analysis of language vis-à-vis earlier language-as-action approaches. Second, I assess the relation between coordinated joint action, which serves as a blueprint for dialogue coordination, and the experience of shared reality, a key concomitant and product of interpersonal communication.

The theory of language production and comprehension (TLPC), laid out in impressive keenness in the Pickering & Garrod (P&G) target article, rests on an analogy between language and action. It musters a throng of cutting-edge research on forward modeling and covert simulation to flesh out the analogy. In the following, I examine the utility of the present action view for the understanding of verbal communication and interpersonal alignment. I will first consider the distinct contribution to the analysis of language vis-à-vis earlier language-as-action approaches. Then I will turn to the relation between coordinated joint action, which serves as a blueprint for dialogue coordination, and a key concomitant and product of interpersonal communication, that is, the experience of a shared reality between interlocutors.

One merit of the TLPC lies in its extension of the action-language analogies that have been championed by a prominent lineage of approaches epitomized by Austin (1962) and Grice (1975). These approaches characterized language use as purposeful contextualized action (Holtgraves 2002). However, much of the research inspired by the language-as-action perspective did not study issues that are now addressed by the TLPC, primarily the online processes that permit the seamless and instantaneous mutual attunement to the current topic (Holtgraves 2002, pp. 180–82). In this respect, the integration of action simulation by the TLPC marks a novel and promising contribution.

However, the focus of the TLPC on co-present interweaving of action, based on prediction processes, outshines key insights from language-as-action work and hence comes at a cost. First of all, the role of context and communicative intentions, essential to any view of language use as action, is mentioned rather parenthetically and relegated to subsidiary information contained in one of the processing modules (viz., the production command). More specifically, it seems that the emphasis on prediction processes underappreciates postdictive processes, that is, the search for an adequate interpretation after utterances have been perceived. Recipients often enough straggle and struggle to infer what is meant by, for instance, nonliteral utterances, figures of speech, or complex (scientific or literary) formulations. According to language-as-action approaches, these efforts are driven by pragmatic assumptions of cooperativeness and mutual adherence to communication rules (Higgins 1981).

Furthermore, the view of action execution and perception, which serves as the blueprint for the meticulous modeling of language production and comprehension, restricts the action-language analogy to co-present, oral dialogue. Given the primacy of face-to-face conversation (e.g., Clark & Wilkes-Gibbs 1986), such a focus of the theory design is reasonable. However, the role of action in other forms of verbal communication remains an open issue. For instance, writers want to accomplish purposes with what they write; readers attempt to identify these purposes. (Processes underlying reading, specifically prediction-by-association, are addressed only once at the end of P&G's article.) To what extent can the action-view embraced by the TLPC account for the processes that operate, for instance, in typing a tweet or blog commentary, and in reading and interpreting it? In reading, there are no immediate sensory perceptions of an interlocutor's movements or the current interaction context that can help a recipient to infer the intended action or purpose underlying a piece of text; but there are other resources, such as assumptions of conversational rules or background knowledge, that allow recipients to make such inferences. It seems that the role of action outside of co-present conversation can be more readily accounted for by other approaches from the "language-as-action" family (for the reception of literary texts, see, e.g., Ricœur 1973).

A second issue I want to examine concerns the relation between coordinated joint action and the experience of a shared reality between the interlocutors. Shared reality is of potentially high relevance to verbal communication because interpersonal communication is a key arena of social sharing. In section 2.3, the models of prediction and simulation are applied to action coordination in joint activities such as ballroom dancing or carrying a bulky object. P&G claim that the joint-action model can explain “the experience of ‘shared reality’ that occurs when A and B realize that they are experiencing the world in similar ways” (target article, sect. 2.3, para. 5).

From the perspective of shared-reality theory, however, the commonality or alignment involved in the coordination of action does not necessarily involve the experience of shared reality. According to a current definition (Echterhoff et al. 2009), shared reality is the product of the motivated process of experiencing an interpersonal commonality of inner states about the world. The creation of a shared reality allows us, for example, to form political or moral convictions, or to evaluate other people or groups. For instance, when people meet a new employee at their workplace, they tend to form shared impressions of the newcomer with their colleagues. As such, and consistent with extant applications of the theory (e.g., Echterhoff et al. 2008), shared reality essentially reflects an *evaluative* alignment between communicator and audience regarding a target entity or state of affairs.

Given this conceptualization, the experience of shared reality and action coordination can be dissociated. Having common representations of an activity and each other’s actions does not mean that the actors have a shared reality regarding how to evaluate the activity. For example, two high-school graduates may perform a seamlessly coordinated ballroom dance at a prom, but the two may feel about and evaluate the dance differently: Whereas one of them may experience it as the exciting beginning of a romantic affair, the other could view it as the mere fulfillment of an obligation, or could even fear subsequent harassment.

This distinction may not come as a surprise because the TLPC is designed to address the key explanandum of psycholinguistic research, that is, the success of conversational interaction, with an emphasis on the rapidity and smoothness of coordination. It is not designed to address higher-level commonalities, such as the sharing of attitudes, evaluations, and judgments. Still, for the sake of further integration and synergy, it may be worthwhile to consider the possible interplay of conversation and shared reality from the perspective of the TLPC.

## The complexity-cost factor in bilingualism

doi:10.1017/S0140525X12002579

Julia Festman

University of Potsdam, Potsdam Research Institute for Multilingualism, 14476 Potsdam, Germany.

festman@uni-potsdam.de

<http://www.uni-potsdam.de/prim/staff/festman.html>

**Abstract:** Language processing changes with the knowledge and use of two languages. The advantage of being bilingual comes at the expense of increased processing demands and processing costs. I suggest considering bilingual complexity including these demands and costs. The proposed model claims effortless monolingual processing. By integrating individual and situational variability, the model would lose its idealistic touch, even for monolinguals.

Because most people today are bilingual, that is, they are language producers and comprehenders of at least two languages, the proposed model by Pickering & Garrod (P&G) cannot be generalized in its current version; it accounts only for a small minority of

language users. Compared to earlier language-processing models, it is more complex—in particular, by integrating production and comprehension—with regard to predicting actions and the embodiment of language. However, it fails to account for the important aspects of increased processing demands and processing costs in bilingual language interaction. These additional demands and costs are observed even for bilinguals with native-like knowledge of the language (so-called *highly proficient bilinguals*) and cannot be attributed to a low level of proficiency. I consider these aspects most relevant for a truly integrated, up-to-date, and generalizable theory of language production and comprehension.

When comparing the performance of monolinguals and bilinguals on the same task, we can trace increased processing demands and processing costs with different measures, two of which I will describe. Such studies investigate the language production in only one language at a time, not during switching between languages and do not make claims about switch costs.

First, increased processing demands for bilinguals can be observed in functional magnetic resonance imaging (fMRI) studies in terms of modulation of the blood-oxygen-level-dependent (BOLD) signal. A recent study (Parker Jones et al. 2011) compared performance of monolinguals and highly proficient bilinguals on naming pictures and reading words aloud. They demonstrated that the same five left hemisphere areas sensitive to increasing demands on speech production in monolinguals showed higher activation in bilinguals. More specifically, during word retrieval and articulation, higher activation was found in dorsal precentral gyrus, pars triangularis, pars opercularis, superior temporal gyrus, and planum temporale. Word retrieval was more demanding for bilinguals than for monolinguals.

Second, processing costs for highly proficient bilinguals are also reported from simple word retrieval tasks. Reaction time latencies are longer for bilinguals than for monolinguals when they were naming pictures or producing a list of tokens from a common semantic category (e.g., Gollan et al. 2002; 2005).

Increased language-processing demands are attributed to the speaker’s necessity of dealing with two languages. Parallel activation of both languages was observed at all stages of bilingual speech production (e.g., Guo & Peng 2006). Consequently, task performance in one of two available languages involves competition of words of the unwanted language with those of the intended language. Therefore, lexical selection is a more demanding process for bilinguals than for monolinguals. As a result, retrieval of words is slower even in highly proficient bilinguals of all ages than in monolinguals (Sorace 2011).

Unbalanced bilinguals are better in one language compared with the other. Hence, they have a stronger and a weaker language. Usually their mother tongue is the stronger language whereas a second, later learned language is the weaker language, characterized by incomplete acquisition of lexicon and grammar. If the first language undergoes attrition (that is, language loss due to infrequent use of that language as a consequence of migration), this language becomes the weaker language and the second language develops to be the stronger language. In both cases, if language proficiency is rather low, it is even more demanding to produce that weak language. It requires more resources to inhibit the stronger, unwanted language (Meuter & Allport 1999). This means that using a weak language involves much language control to avoid unintentional switching to the stronger language. Several processing models for bilingualism (e.g., Green 1986) capitalize on the concept of language control and emphasize the problem of limited resource and processing capacities. Applying the proposed theory to a communicative setting with unbalanced bilinguals, I seriously question, in particular, the suggested obligatory action predictions and the effortless involvement in interactive language.

P&G acknowledge one condition under which predictions of one’s own and other’s actions can be difficult, namely, if they are “unrelated” (an example for a related action would be

ballroom dancing; see target article, sect. 2.3, Joint Action). In my view, in the domain of language, the ease of prediction depends not only on relatedness, but also, and much more, on familiarity. Not only are predictions easier and more likely to be correct when communicative participants are familiar with each other and their communicative habits, but also the interlocutors must be familiar with the language they are using for the interaction, and with the culture of that language. It is certainly rather easy to predict actions of communication in a long-standing relationship (e.g., the best friend), but for an exchange student – age 16, foreign to a country, new to a host family, and with little knowledge either of the language spoken or of the family's speech habits – the success rate of other's action prediction is most probably low. Similarly, in foreign-language reading, cultural familiarity has been found to be crucial for comprehension. Studies on non-native reading revealed that if readers lack the relevant cultural knowledge, reading activities could not fully compensate for the discrepancy or help readers comprehend a text (Erten & Razi 2009).

With all this in mind, it becomes clear that the language producer in the proposed model is ideal, even as a monolingual. There are moments when we suffer a snub from another person's action: we are so surprised that words fail us. Some seem to have the gift of gab, and are more quick-witted than others. For such situational or inter-individual variability known from everyday life, the current model does not account. It holds the view that interlocutors are perfectly coordinated and that there is no time gap between interlocutors' turns. Nonetheless, P&G have been aware of differences between native and non-native speakers or of the difference between speaking to an adult or a child (see sect. 4, General Discussion) to some degree, but they did not integrate these differences into their model. The authors might want to elaborate on situations and conditions when the construction of forward models, covert imitations, joint actions, and continuous prediction generation are not ideal, not even for the monolingual.

## An ecological alternative to a “sad response”: Public language use transcends the boundaries of the skin

doi:10.1017/S0140525X12002580

Carol A. Fowler

Department of Psychology, University of Connecticut, Storrs, CT 06269.

[carol.fowler@uconn.edu](mailto:carol.fowler@uconn.edu)

<http://web2.uconn.edu/psychology/people/Faculty/Fowler/Fowler.html>

**Abstract:** Embedding theories of language production and comprehension in theories of action-perception is realistic and highlights that production and comprehension processes are interleaved. However, layers of internal models that repeatedly predict future linguistic actions and perceptions are implausible. I sketch an ecological alternative whereby perceiver/actors are modeled as dynamical systems coupled to one another and to the environment.

In investigating between-person language use, Pickering & Garrod (P&G) have taken a road less traveled. Psychology of language has been dominated by studies of private language processing within individuals, respecting the view of Noam Chomsky that public language, “however construed, appears to have no significance” (1986, p. 31). I particularly admire Garrod and Pickering's (2004) paper in which they invoked between-person alignments at multiple linguistic levels to explain why, for most of us, conversation is easier than monologue even though conversation requires coordination with others.

The approach presented in the target article is important and valuable in other ways. One is the recognition that language production and comprehension at its various descriptive levels are

species of action and perception that, accordingly, require explanations consistent with those of nonlinguistic acting and perceiving. A second is the recognition that, in contrast to typical theoretical treatments, language production and comprehension are thoroughly interleaved. To reflect that, Pickering & Garrod (P&G) weaken the “horizontal split” in their Figure 1 that separates processes of comprehension and production within an individual.

However, in my view, the particular integration of production and comprehension processes that the authors propose is unrealistic in consisting of a complex cognitive tiling of predictive modeling processes (that is, predictions at multiple levels that occur repeatedly over time). The proposal falls into the category of theoretical accounts that Bentley (1941) identified as “sad responses,” that *are* sad in unrealistically ascribing responsibility for behavioral systematicities in the world almost exclusively to processes “in the head” or “in the brain” (p. 13).

I recommend a different route to understanding language use that involves weakening separations not only along the horizontal dimension of P&G's Figure 1, but also along the vertical dimension, the one that separates entities bounded, as Bentley (1941) put it, by the skin. Warren (2006) offered an integrated theory of perception and action along these lines, and Marsh et al. (e.g., Marsh et al. 2006) provided a compatible ecological approach that encompasses social interactions.

Warren (2006) explicitly rejected the model-based approaches adopted by P&G on grounds that, (a) in them, implausibly, actor/perceivers are supposed to interact directly with their internal models, but indirectly with the world itself, and (b) interacting with the models means using representations whose origins must appeal “in circular fashion to the very perception and action abilities they purport to explain” (p. 361). He offered instead an approach in which perceiver/actors and their environments are modeled as dynamical systems that are coupled both mechanically and informationally. Adaptive behavior emerges from constraints arising from the structure of the environment, the biomechanics of the body, perceptual information about the coupled system, and task demands. In the approach of Marsh et al. to social perception and action (e.g., Marsh et al. 2006; 2009), the coupled dynamical systems relevant to understanding social activity include more than one perceiver/actor. In the context of this account, the kinds of alignments among interlocutors that Garrod and Pickering (2004) identified as fostering successful conversation are reflections of the interpersonal coordinations that occur when humans come together to engage in joint activities.

This approach has promise for understanding interpersonal language use in two ways that are relevant to themes in the target article.

One is that the approach promises to obviate postulating the multiple levels of repeated predictive modeling that characterizes P&G's account. Some perceptual information is prospective in nature in signaling what will happen. Moreover, use of prospective stimulus information (e.g., about nutrients at some distance) occurs among organisms that lack a nervous system and hence lack the means to construct forward and inverse models (see, e.g., Reed 1996). Turning to humans, an outfielder for example, does not need to predict where and when a fly ball will become catchable; structure in reflected light over time provides prospective information about the ball's future trajectory (e.g., Michaels & Oudejans 1992). Likewise, information in reflected light can signal whether it is safe to cross a street in traffic without pedestrians having to predict whether or when a vehicle will cross their path (e.g., Oudejans et al. 1996). In short, prospective information can constrain action without intervening predictions being made. In language, the prospective information can be pragmatic, syntactic, lexical, semantic, phonological, and so on.

A second domain in which the ecological approach may have promise for understanding language production and comprehension concerns the pervasive findings to which P&G allude of

imitation or, less often, complementation in language use and elsewhere. I suggest that imitation occurs pervasively in laboratory research, because perception of the actions of someone else essentially serves as instructions for imitation (e.g., Fowler et al. 2003). Therefore, it is often easy to imitate, and imitation can be a default reflection of the disposition of humans to coordinate with one another (e.g., Richardson et al. 2007). In general, however, imitation is not always an adaptive response to the actions of someone else. Nor is complementing what is perceived. Missing from much research on “embodied cognition” are task constraints that encourage anything other than default mirroring. However, when imitation is discouraged, embodied responses to language input may occur that are not imitative (cf. Olmstead et al. 2009). In the context of tasks that make other kinds of interpersonal coordinations relevant, imitative responses may not be as pervasive as they are in typical laboratory research.

Indeed, an important issue for understanding public language concerns how in general utterances constrain users’ behavior. When someone shouts *Duck!* to a bicyclist obliviously approaching a low-hanging branch, an adaptive response is, in fact, to duck (not to imitate the utterance). Research using the visual world procedure (e.g., Ferreira & Tanenhaus 2007), albeit designed with other purposes in mind, suggests that language can affect nonimitative adaptive action (in this case, by the eyes) very quickly. Investigation of language use embedded in meaningful contexts may help to reveal whether or not imitation is fundamental.

## Are forward models enough to explain self-monitoring? Insights from patients and eye movements

doi:10.1017/S0140525X12002749

Robert J. Hartsuiker

Department of Experimental Psychology, Ghent University, B-9000 Ghent, Belgium.

robert.hartsuiker@ugent.be

<http://users.ugent.be/~rhartsui/>

**Abstract:** At the core of Pickering & Garrod’s (P&G’s) theory is a monitor that uses forward models. I argue that this account is challenged by neuropsychological findings and visual world eye-tracking data and that it has two conceptual problems. I propose that conflict monitoring avoids these issues and should be considered a promising alternative to perceptual loop and forward modeling theories.

At the core of Pickering & Garrod’s (P&G’s) theory of language production and comprehension is the monitor, a cognitive mechanism that allows speakers to detect problems in speech. The proposal is that the monitor compares perceptual representations from two channels, namely (a) a perceptual representation of the semantics, syntax, and phonology that the “production implementer” is producing; and (b) forward perceptual representations of forward production representations of each linguistic level. If there is a mismatch, the monitor has detected the problem and so can begin a correction. Monitoring via forward models is part of an elegant account that nicely integrates the action and language literatures. But is this account compatible with findings on speech monitoring?

The model shares with perceptual loop theories (Hartsuiker & Kolk 2001; Levelt 1989) the assumption that monitoring needs speech comprehension (i.e., to create a perceptual representation of produced speech). It therefore shares the problems of other models with a perceptual component. One problem is that neuropsychological studies found dissociations between comprehension and monitoring in brain-damaged patients. A striking example is a 62-year-old woman with auditory agnosia and

aphasia reported by Marshall et al. (1985). Although this patient’s auditory system was intact, she was unable to comprehend familiar sounds, words, or sentences and could not report number of syllables or stress contrasts. Speech production was seriously impaired, with speech often containing neologistic jargon. But despite the patient’s severe comprehension problems, she produced a great many (often unsuccessful) attempts at self-corrections of errors—in particular, her phonological errors (but not her semantic errors). These findings suggest that error detection can take place without perception.

Another patient that challenges perception-based monitoring is G., a 71-year-old Dutch Broca’s aphasic (Oomen et al. 2005). In a speech production task, G. produced many phonological errors, of which he repaired very few, whereas he produced very few semantic errors, which he usually repaired. G.’s production difficulties hence mirror his monitoring difficulties. Importantly, G.’s difficulty to repair phonological errors cannot be attributed to a perception deficit: In a perception task, he detected as many phonological errors as a group of controls. It therefore seems that G.’s monitoring deficit is related to this production deficit and not to any perception deficit. These data argue against a forward model account, because the forward model is a separate and qualitatively different system from the production implementer. There is no reason why the monitoring deficit should mirror the production deficit.

Our recent visual world eye-tracking data (Huetig & Hartsuiker 2010) also speak against both perceptual loop and forward modeling accounts. When our subjects named a picture of a *heart*, they gazed at the phonologically related written word *harp* more often than at unrelated words, and this “competitor effect” had the same time course as the analogous effect when listening to someone else (Huetig & McQueen 2007). These data are not consistent with perceptual loop accounts, which predict earlier competitor effects in production than in comprehension, because the phonological representation inspected by the inner loop precedes external speech by a considerable amount of time (namely, the time articulation takes). Similarly, the representations created by forward models also precede overt speech in time, so a monitoring with forward models account also predicts an early competitor effect. After all, forward models are *predictions* of what one will say, and fixation patterns in the visual world are strongly affected by prediction. One might object that the predicted phonological percept is too impoverished to create a phonological competitor effect. But this seems to contrast with Heinks-Maldonado et al.’s (2006) magnetoencephalography data showing reafference cancellation with frequency-shifted feedback, which implies a highly detailed phonological percept that even includes pitch.

There are also conceptual issues with monitoring via forward models. One is the reduplication of processing systems (Levelt 1989). Specifically, the production implementer creates semantic, syntactic, and phonological representations whereas a forward model creates corresponding representations. If the forward model creates highly accurate representations at each level, we have two separate systems doing almost the same thing, which is not parsimonious. But if the forward model creates highly impoverished representations (e.g., only one phoneme), such representations are not a good standard for judging correctness. It then becomes difficult to see how speakers detect so many errors at so many linguistic levels.

Additionally, if we assume the output of forward models, although impoverished, is still good enough to be useful for the monitor, then the monitor will have to “trust” the forward model, just like P&G’s metaphorical sailor having to trust his charted route. But there is no *a priori* reason for assuming that the forward model is less error prone than the production implementer; in fact, if forward models are “quick and dirty” they will be *more* error prone. Trust in the forward model will then be misplaced. But such misplaced trust has the undesirable effect of creating “false alarms,” so that a correct item is replaced by an

error. “Corrections” that make speech *worse* do not seem to occur frequently, although some repetitions may in fact be misplaced corrections (e.g., Hartsuiker & Notebaert 2010).

P&G briefly mention an alternative to both the perceptual loop account and the forward model account, namely, a conflict monitoring account (Botvinick et al. 2001; Mattson & Baars 1992; Nozari et al. 2011). According to such accounts, monitoring does not use comprehension, but measures the amount of “conflict” in each layer of production representations, assuming that conflict is a sign of error. Conflict monitoring has the advantages of allowing error detection without perception and that a production deficit at a given level is straightforwardly related to an error detection deficit at that level. Such an account is consistent with Huettig and Hartsuiker’s (2010) eye-tracking data and avoids the reduplication of processing components. Finally, in a conflict monitoring account, there is also no issue of which representation to “trust.” It is worth therefore considering conflict monitoring as a viable alternative to the perceptual loop and forward modeling accounts.

## Predictive coding? Yes, but from what source?

doi:10.1017/S0140525X12002750

Gregory Hickok

Department of Cognitive Sciences, University of California, Irvine, CA 92697.  
gshickok@uci.edu  
<http://alns.ss.uci.edu>

**Abstract:** There is little doubt that predictive coding is an important mechanism in language processing—indeed, in information processing generally. However, it is less clear whether the action system is the source of such predictions during perception. Here I summarize the computational problem with motor prediction for perceptual processes and argue instead for a dual-stream model of predictive coding.

Predictive coding is in vogue in cognitive neuroscience, probably for good reason, and we are no strangers to the idea in the domain of speech (Hickok et al. 2011; van Wassenhove et al. 2005). The current trendsetters in predictive coding are the motor control crowd who have developed, empirically validated, and promoted the notion of internal forward models as a neural mechanism necessary for smooth, efficient motor control (Kawato, 1999; Shadmehr et al. 2010; Wolpert et al. 1995). But the basic idea has been pervasive in cognitive science for decades in the form of theoretical proposals like analysis-by-synthesis (Stevens & Halle 1967) and in the form of empirical observations like priming, context and top-down effects, and the like. So Pickering & Garrod’s (P&G’s) claim that language comprehension involves prediction is nothing new. Nor is it a particularly novel claim, right or wrong, that the motor system might be involved in receptive language; it has gotten much attention in the domain of speech perception/phonemic processing, for example (Hickok et al. 2011; Rauschecker & Scott 2009; Sams et al. 2005; van Wassenhove et al. 2005; Wilson & Iacoboni 2006) and has been a component of at least some aspects of sentence processing models for decades (Crain & Fodor 1985; Frazier & Flores d’Arcais 1989; Gibson & Hickok 1993). What appears to be new here is the idea that prediction at the syntactic and semantic levels can come out of the action system rather than being part of a purely perceptual mechanism.

This is an interesting idea worth investigation, but it is important to note that there are computational reasons why motor prediction generally is an inefficient, or even maladaptive, source for predictive coding during receptive functions. Here’s the heart of the problem. The computational goal of a motor prediction in the context of action control is to increase perceptual sensitivity to *deviations* from prediction (because something is wrong and correction is needed) and to decrease sensitivity to accurate

predictions (all is well, carry on). Hence, if motor prediction were used in the context of perception, it would tend to suppress sensitivity to that which is predicted, whereas an efficient mechanism should enhance perception. Behavioral evidence bears this out. The system is less sensitive to the perceptual effects of self-generated actions (unless there is a deviation) than to externally generated perceptual events. Some of the “reafference cancellation” effects noted by P&G are good examples: inability to self-tickle, saccadic suppression of motion percepts, and the motor-induced suppression effect measured electrophysiologically. This contrasts with nonmotor forms of prediction, what P&G referred to as the association route, which might include context effects and priming and which tend to facilitate perceptual recognition. Put simply, motor prediction decreases perceptual sensitivity to the predicted sensory event, nonmotor prediction increases perceptual sensitivity to the predicted sensory event. Why, then, is there so much attention on motor-based prediction?

P&G argue that there is evidence to support a role for motor prediction in language-related perceptual processes—a good reason to focus attention on a motor-based prediction process. There are problems with the evidence they cite, however. One cannot infer causation from motor activation during perception (it could be pure associative priming [Heyes, 2010; Hickok, 2009a]), the transcranial magnetic stimulation (TMS) evidence in speech perception tasks is likely a response bias effect (Venezia et al. 2012), and the studies showing effects of imitation *training* on perception do not necessarily imply that imitation is carried out *during* perception, which is the claim that P&G wish to make.

There is a better way to conceptualize the architecture of the system, one that flows naturally out of fairly well-established models of cortical organization (Hickok & Poeppel 2007; Milner & Goodale 1995). A dorsal stream subserves sensory-motor integration for motor control; it is a highly adaptable system (Catmur et al. 2007) that links sensory targets (objects in space, sequences of phonemes) with motor systems tuned to hit those targets under varying conditions. A ventral stream subserves the linkage between sensory inputs and conceptual memory systems; it is a more stable system designed to abstract over irrelevant sensory details. Both systems enlist predictive coding as a fundamental computational strategy (Friston et al. 2010), but both in the service of what the systems are designed for computationally. Motor prediction facilitates motor behavioral (but suppresses perception) and “sensory” or “ventral stream” prediction facilitates perception (Hickok 2012b).

P&G underline that their approach blurs the line between comprehension and production and thus rejects the “cognitive sandwich” view, whereas the alternative perspective just outlined might be interpreted as preserving the comprehension-production distinction. In this context, it is worth pointing out that P&G do not actually blur the distinction between the two slices of bread all that much. They are quite distinct computational and representational components as their *c* and *p* notation attests, and they have even added some slices, an action implementation system (*p*), a forward production model (*p-hat*), a forward comprehension model (*c-hat*) and a perceptual system (*c*), each of which generates phonological, syntactic, and semantic representations—nearly a loaf of bread. They do argue, correctly in my view, and consistent with many speech scientists and motor control researchers as well as the classical aphasiologists (despite P&G’s claims to the contrary), that comprehension and production systems must interact. We make the same claims of our dorsal stream (Hickok 2012a; Hickok & Poeppel 2007). But where P&G and others—including myself (Hickok et al. 2011)—have gone wrong, in my view, is that they are trying to shoehorn a motor-control-based mechanism into a perceptual system that it was not designed to serve.

## ACKNOWLEDGMENT

Supported by NIH grant DC009659.

## “Well, that’s one way”: Interactivity in parsing and production

doi:10.1017/S0140525X12002592

Christine Howes, Patrick G. T. Healey, Arash Eshghi, and Julian Hough

Queen Mary University of London, Cognitive Science Research Group,  
School of Electronic Engineering and Computer Science, London E1 4NS,  
United Kingdom.

c.howes@qmul.ac.uk ph@eecs.qmul.ac.uk  
arash@eecs.qmul.ac.uk julian.hough@eecs.qmul.ac.uk  
<http://www.eecs.qmul.ac.uk/~chrisba/>

**Abstract:** We present empirical evidence from dialogue that challenges some of the key assumptions in the Pickering & Garrod (P&G) model of speaker-hearer coordination in dialogue. The P&G model also invokes an unnecessarily complex set of mechanisms. We show that a computational implementation, currently in development and based on a simpler model, can account for more of this type of dialogue data.

Pickering & Garrod’s (P&G’s) programmatic aim is to develop an integrated model of production and comprehension that can explain intra-individual and inter-individual language processing (Pickering & Garrod 2004; 2007). The mechanism they propose, built on an analogy to neuro-computational theories of hand movements, involves producing and comparing two representations of each utterance; a full one containing all the structure necessary to produce the utterance and an “impoverished” efference copy that can predict the approximate shape the utterance should have.

Although not our central concern, there is a tension between endowing the efference copy with enough structure to be able to predict semantic, syntactic, and phonetic features of an utterance and nonetheless making it reduced enough that it can be produced ahead of the utterance itself. To avoid a situation in which the “impoverishment” proposed for the efference copy is just those things not required to fit the data, we need independently motivated constraints on its structure.

Neuro-computational considerations might provide such constraints, but there are dis-analogies with the models of motor control P&G use as motivation. Efferent copies were originally proposed to enable rapid cancellation of self-produced sensory feedback, for example, to maintain a stable retinal image by cancelling out changes due to eye-movements. However, the claim that we use an analogous mechanism to predict, and correct, linguistic structure before an utterance is produced involves something conceptually different. The awkwardness of phrases such as “semantic percept” highlight this difference; until the utterance is actually produced there is nothing to generate the appropriate sensory percept. Conversely, if the “percept” is internal we are still in the cognitive sandwich.

These points aside, the target article provides a valuable overview of the evidence that language production and comprehension are tightly interwoven. P&G’s main target, the “traditional model”, treats whole sentences, “messages” or utterances as the basic unit of production and comprehension. However, there is evidence from cognitive psycholinguistics and neuroscience to show that language processing is tightly interleaved around smaller units. The close interconnections between production and comprehension are especially clear in dialogue where fragmentary utterances are commonplace and people often actively collaborate with each other in the production of each turn (Goodwin 1979).

It is unclear if the interleaving of production and comprehension requires internally structured predictive models. Recent progress on incremental models of dialogue suggest a more parsimonious approach. In our computational implementation based on Dynamic Syntax (Purver et al. 2006; 2011; Hough 2011), the burden of predicting full utterances does not need to be employed in parsing, as speakers and hearers have incremental access to representations of utterances as these emerge. Contrarily, P&G’s approach to self-repairs is analogous to Skantze and

Hjalmarsson’s (2010), which compares string-based plans and computes the difference between the input speech plan and the current state of realisation. In our model, instead of having to regenerate a new speech plan from scratch, we can repair the necessary increments, reusing representations already built up in context, which are accessible to both speaker and hearer. Currently, it is difficult to distinguish empirically between a dual-path model with predictions and a single-path incremental model because both combine production and comprehension.

As the paper highlights, the “vertical” issue of interleaving production and comprehension is independent from the “horizontal” problem of accounting for how language use is coordinated in dialogue. Nonetheless, this article extends previous Pickering and Garrod work (2004; 2007) in claiming that the model of intra-individual processing can be extended to inter-individual language processing (conversation). Unlike previous work, the new model operates in different ways for speakers and hearers, and the potential for differences between people’s dialogue contexts is acknowledged (although not directly modelled).

The problem with this generalisation is that in dialogue we do not just predict what people are going to say, we also respond. Even if I could predict what question you are about to ask, this does not determine my answer (although it might allow me to respond more quickly). In terms of turn structure, all a prediction can do is make it easier for me to repeat you. Repetition does occur in dialogue but is rare and limited to special contexts. Corpus studies (Healey et al. 2010) indicate that we repeat few words (less than 4%) and little more syntactic structure (less than 1%) than would be expected by chance. Crudely, a cross-person prediction model of production-comprehension cannot explain 96% of what is actually said in ordinary conversation.

One conversational context that seems to depend on the ability to make online predictions about what someone is about to say is compound contributions, in which one dialogue contribution continues another, as in this excerpt from Lerner (1991):

Daughter: Oh here Dad, one way to get those corners out

Father: is to stick your fingers inside

Daughter: Well, that’s one way.

Although it is unclear whether a predictive model better accounts for the father’s continuation than one in which he is building a response based on his partial parse of the linguistic input, the daughter’s response seems to be based on the mismatch between what was said and what she had planned to say. Although possible she was predicting he would say what she herself had planned to, there is no need for this additional assumption. Many cases of other-repair (Schegloff 1992) such as clarification requests asking what was meant by what was said (e.g., “what?”) also seem to require that any predictability used is impoverished at precisely the level it might be useful.

In a study on responses to incomplete utterances in dialogue (Howes et al. 2012), increased syntactic predictability led to more clarification requests. Although participants made use of different types of predictability in producing continuations, predictability was neither necessary nor sufficient to prompt completion, and, in extremely predictable cases, participants did not complete the utterance, responding as if the predictable elements had been produced. Our assumption is that it is the things we cannot predict that are the most important parts of conversation. Otherwise, it is hard to see why we should speak at all.

## Seeking predictions from a predictive framework

doi:10.1017/S0140525X12002762

T. Florian Jaeger<sup>a,b</sup> and Victor Ferreira<sup>c</sup>

<sup>a</sup>Department of Brain and Cognitive Sciences, University of Rochester,  
Rochester, NY 14627-0268; <sup>b</sup>Department of Computer Science, University of

Rochester, Rochester, NY 14627; <sup>°</sup>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; Sonderogger & 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 morpho-syntactic 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 efferece 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 efferece 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 efferece 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,<sup>a</sup> Nicholas B. Turk-Browne,<sup>a</sup> and Adele E. Goldberg<sup>b</sup>

<sup>a</sup>Department of Psychology, Princeton University, Princeton, NJ 08544;

<sup>b</sup>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 the 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-NP<sub>theme</sub>-NP<sub>locative</sub>* (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-NP<sub>theme</sub>-NP<sub>locative</sub>* 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,<sup>a</sup> Harold Bekkering,<sup>b</sup> and Emiko S. Kashima<sup>c</sup>

<sup>a</sup>Melbourne School of Psychological Sciences, The University of Melbourne, Parkville, Victoria 3010, Australia; <sup>b</sup>Donders Institute for Brain, Cognition, and Behaviour, Radboud University Nijmegen, 6500 HE Nijmegen, The Netherlands; <sup>c</sup>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

that modulates its operation in the service of the language users' communicative intentions. Implications for intergroup relationships and intercultural communication are discussed.

Pickering & Garrod (P&G) collapse the oft-made separation between comprehension and production, and propose to reconceptualise language processing in terms analogous to the ideomotor account of action perception and motor action. In a nutshell, the ideomotor principle (1) rejects the modular view of perception, cognition, and action, dubbed by Hurley (2008a) as the cognitive sandwich, and (2) suggests that perception and action are closely linked together, and cognition acts as a fallible control mechanism that modulates and is modulated by the dynamic perception-action link. P&G highlight the perception-action link and reconceptualise language comprehension ( $\approx$ perception) and production ( $\approx$ action) as highly dynamic and closely linked processes that cannot be separated. In so doing, their proposal sheds an intriguing light on the mystery of language-based social interaction – how humans can communicate with each other so efficiently and smoothly to carry out their joint activity despite some obvious issues of occasional miscommunication and communication breakdown.

We applaud their attempt to ground language use in the mechanisms of motor perception and action, and we endorse their theorizing that the perception-action substrate provides an “infrastructure” of language-based social interaction. However, we also believe that this infrastructure needs to be complemented by an additional control mechanism that modulates its operation, or *cognition*, the “ideo” part of ideomotor theory. We suggest that one such mechanism is language users' communicative intentions.

**What are communicative intentions?** Language is typically used within the social context of joint activities (e.g., Clark 1996; Kashima & Lan, in press). As Bratman (1992) noted, commitment to a joint activity and readiness to be responsive to the intentions and actions of one's partner are integral to an intentional joint activity. The participants in a joint activity hold joint intentions, or intentions to carry out their individual parts of the joint activity to attain their joint goal while coordinating each other's actions (e.g., Bratman 1999; Pacherie 2012; Tuomela 2007). By communicative intentions we mean intentions to carry out largely the linguistic part of the joint activity by performing illocutionary and locutionary acts to attain its goal.

Indeed, P&G briefly alluded to the importance of joint intentions in language use: “linguistic joint action is more likely to be successful and well-coordinated than many other forms of joint action, precisely because the interlocutors communicate with each other and share the goal of mutual understanding” (target article, sect. 3.3, para. 8). Echoing their sentiment and building on it more explicitly, we argue that intentional communicative mechanisms can partly regulate the comprehension-production processes, and this can have significant sociocultural implications.

**Intentions and perception-action link.** The modulation of perception-action link by shared intentions was suggested by Ondobaka et al.'s (2011) work. In this experiment involving a confederate and a naïve participant, the confederate first performed an action (e.g., selecting the higher of two numbers) on each trial, and the participant performed an action congruent or incongruent with the confederate's action intention, that is, doing the same thing (selecting a higher number) or the opposite (selecting a lower number). However, the correct action for the participant was motorically congruous with the confederate's action in some cases (selecting the number displayed on the same side), but incongruous in others (selecting the number on the opposite side). If the perception-action link is fixed and not intentionally modulated, the participant's perception of an action should facilitate a motorically congruent action, but interfere with a motorically incongruent action regardless of action intentions. On the contrary, this occurred *only when* action intentions were congruous (i.e., the co-actors shared the same goal). When action intentions were incongruous, congruity of motor actions

had no effect. These results are consistent with Carpenter's (1852) original proposal of the ideomotor principle, which goes beyond perception-guided movement and stresses the importance of conceptual action intention and expectation in the guidance of voluntary behaviour (see Ondobaka & Bekkering 2012, for a recent review).

Analogously, communicative intentions may modulate the linguistic perception-action link. For instance, conversants' accents often converge (e.g., Giles et al. 1991), and this can be interpreted within P&G's framework. However, there is evidence to suggest that accent convergence depends on the conversant's communicative intentions. Bourhis and Giles (1977) examined bilingual English-Welsh speakers' accent change in a conversation with an English interviewer. When their Welsh identity was threatened, their accent depended on their personal goals. Those who were learning Welsh to extend their careers showed a convergence to the interviewer's accent, whereas those who were learning the language because of their Welsh identity showed a *divergence*, strengthening their Welsh accent. Babel's (2010) recent study partly replicated this finding. Native speakers of New Zealand English listened to an Australian English speaker and pronounced the same words (i.e., shadowed productions). Despite the similarities between New Zealand and Australian English dialects, there are detectable and systematic differences. The participants who had more-negative implicit attitudes towards Australia (vs. New Zealand) showed less convergence to the Australian accent.

These findings in language use as well as nonverbal mimicry (e.g., Castelli et al. 2009) suggest that communicative and action intentions can modulate the verbal and nonverbal perception-action link in joint activities. This raises at least two classes of critical questions. First, *how* do communicative and action intentions regulate the embodied substrate of social interaction? What are the mechanisms for the dynamic and mutual influences between the conceptual strata of intentions and the embodied strata of perception-action link? Second, what social and cultural processes contribute to the shaping of the two strata, which in turn can have a long-term impact on the evolution of macro structures such as culture and society (e.g., Holtgraves & Kashima 2008)? In particular, the interaction of intentions and embodiment may play a critical role in intergroup differentiation (e.g., Welsh vs. English; Australian vs. New Zealander), the maintenance and dissolution of the intergroup boundary, as well as intergroup and intercultural communication and understanding.

## Preparing to be punched: Prediction may not always require inference of intentions

doi:10.1017/S0140525X12002622

Helene Kreysa

Department for General Psychology and Cognitive Neuroscience, Friedrich Schiller University Jena, 07743 Jena, Germany.

helene.kreysa@uni-jena.de

<http://www2.uni-jena.de/svw/allgpsy/team/kreysa-h.htm>

**Abstract:** Pickering & Garrod's (P&G's) framework assumes an efference copy based on the interlocutor's intentions. Yet, elaborate attribution of intentions may not always be necessary for online prediction. Instead, contextual cues such as speaker gaze can provide similar information with a lower demand on processing resources.

At several points in their target article, Pickering & Garrod (P&G) suggest that prediction by simulation is based on determining a conversational partner's intention through perceiving the unfolding action or speech, potentially combined with background knowledge. On this basis, an efference copy of the intended act is generated, enabling the prediction of upcoming behavior

and the production of behavior or speech that complements it. But how explicit do the attributed intentions need to be to allow such prediction, and how might comprehenders derive them?

Unfortunately, P&G do not define clearly what they mean by “intention”: Whereas on the part of the actor or speaker, an intention seems to represent nothing more than an “action command” (e.g., in the legend to Figure 3; target article, sect. 2.2), recognizing intentions on the part of the comprehender is more complicated. According to P&G, it can involve considerations of past and present behavior and of the speaker’s perceived state of mind, as well as ongoing modifications of this interpretation. In the literature, identifying others’ intentions is generally taken to imply an additional step beyond action recognition, that of identifying the goal of this action (see e.g., Levinson 2006; Tomasello et al. 2005). A similar view underlies the HMOSAIC architecture (Wolpert et al. 2003), which contains symbolic representations of the task in the form of goals or intentions.

Elaborate attributions of intention based on the interlocutor’s potential motivations in the current situation and on general world knowledge are certainly useful for understanding speech. They help to generate expectations about how the conversation is likely to develop and can be used for what P&G call “offline prediction.” Yet although such expectations have, in turn, been shown to influence moment-by-moment language comprehension (e.g., Kamide et al. 2003; Van Berkum et al. 2008), they are presumably not computed on a moment-by-moment basis and remain relatively constant across extended periods of time.

The time-critical online simulations in P&G’s account of real-time conversation must require intentions of a more basic kind—some form of heuristic for anticipating others’ upcoming actions. To use their example: It is unquestionably valuable if I am quick to predict that someone is preparing to punch me rather than to shake my hand, so I can prepare to move appropriately. But if I start considering why they might wish to hurt me, I will probably be too late in responding.

For real-time online prediction of upcoming words and sentences, I would like to suggest that it may often be possible to rely on contextual clues to a speaker’s upcoming actions that are directly perceivable: A particular tone of voice, a facial expression, a hand gesture, or a shift in the speaker’s gaze direction can all be informative about how the speaker plans to continue a sentence. In this sense, such cues are closely connected to the speaker’s intentions. At the same time, they are often produced unintentionally on the part of the speaker, and can be readily perceived on the part of the comprehender. This is exactly what makes them so efficient: They can help to disambiguate the linguistic signal without requiring deliberate consideration of intentions (cf. Shintel & Keysar 2009, who refer to such processes as “non-strategic generic-listener adaptations”; p. 269).

A prime example of this type of contextual cue is the direction of other people’s gaze. Gaze is a salient attentional cue that reliably causes viewers to shift their own attention in the same direction (Emery 2000). This occurs even in the visual periphery and without requiring conscious awareness (Langton et al. 2000; Xu et al. 2011). Additionally, because speakers tend to look at objects they are preparing to mention (e.g., Griffin & Bock 2000; Meyer et al. 1998), gaze will often directly reflect the speaker’s action plan. Such referential gaze is therefore both easy to detect and informative about upcoming sentence content. In fact, comprehenders can and do make rapid use of the speaker’s gaze direction to anticipate upcoming referents (Hanna & Brennan 2007; Staudte & Crocker, 2011) and even to assign thematic role relations (Nappa et al. 2009; Knoeferle & Kreysa 2012).

These benefits of gaze-following in comprehension can be conceived of as a form of prediction about what will be mentioned next, similar to anticipatory fixations of objects in the visual world in eye tracking studies of spoken sentence processing (e.g., Altmann & Kamide 1999; Knoeferle & Crocker 2006; for

recent reviews see Altmann 2011 and Huettig et al. 2011). It is interesting to consider P&G’s classification and to speculate on whether this might be *prediction by association* (e.g., “people who look at objects often mention them shortly thereafter”) or even *prediction by simulation* (using one’s own gaze behavior as a proxy, e.g., “if I had just looked at the kite, I’d refer to it next”). Alternatively, it might be sufficient that speaker gaze attracts the comprehender’s attention to a location that is relevant for understanding the unfolding speech utterance. In all three cases, the end result is a coordination of the interlocutors’ attention on the same objects in the visual world. This is known to benefit problem solving (Grant & Spivey 2003; Knoblich et al. 2005) and conversation in general (Richardson & Dale 2005; Richardson et al. 2007). Such benefits may well be due to a shared perspective and aligned representations of the situation, but they need not imply awareness of the interlocutor’s intentions.

## A developmental perspective on the integration of language production and comprehension

doi:10.1017/S0140525X12002774

Saloni Krishnan

Centre for Brain & Cognitive Development, Birkbeck, University of London, London WC1E 7HX, United Kingdom.

s.krishnan@psychology.bbk.ac.uk

<http://www.cbcd.bbk.ac.uk/people/students/Saloni>

**Abstract:** The integration of language production and comprehension processes may be more specific in terms of developmental timing than Pickering & Garrod (P&G) discuss in their target article. Developmental studies do reveal links between production and comprehension, but also demonstrate that the integration of these skills changes over time. Production-comprehension links occur within specific language skills and specific time windows.

Pickering & Garrod (P&G) have set out an argument for a possible integration of language production and comprehension processes in adults. However, charting how these two processes come to be set up during development is critical in attempts to understand their subsequent integration. An obvious problem when considering production/comprehension processes within a developmental framework is that language production lags behind language comprehension. So, to reiterate P&G’s question, can silent naming use the production system, if no word production has yet emerged? Furthermore, the dynamic and nonmonotonic changes in the development of children’s motor systems, as well as the specificity of motor-language links during these stages, need to be integrated into any such model. The need to examine developmental evidence in building this theory is therefore critical.

Numerous links between production and comprehension occur during language learning in childhood. These links have been reported in developmental studies assessing gesture production and language comprehension (Bates & Dick 2002; Iverson & Thelen 1999). From an atypical development standpoint, there is also a greater incidence of co-morbid motor coordination and planning difficulties in children with language impairment (Iverson & Braddock 2011). In support of P&G’s model, there are studies specifically demonstrating how auditory perception/language comprehension can also affect speech motor performance in childhood. For instance, perceptual ability can influence the learning of motor gestures. Seemingly due to the complexity of articulation, affricates are produced later in development for English-speaking children. By contrast, for Putonghua-speaking children, affricates are acquired very early, probably due to their salience within the language (Dodd & McIntosh 2010). Furthermore, for higher cognitive-linguistic demands as

compared to lower ones, speech motor variability also increases (reviewed in Goffman 2010), indicating that the production processes are influenced by comprehension/perceptual processes. Indeed, a catalyst to changes in motor control may be vocabulary increases (Green & Nip 2010). Speech motor variability can also act as an index of learning; kinematic analyses of motor movements reveal that children receiving training for articulatory disorders produce different motor gestures associated with phonetic categories, which were imperceptible at the acoustic level (Gibbon 1999).

However, it is important to note that these links do not extend to all motor skills, and in particular, not to gross motor ability as assessed by locomotion or play (Bates et al. 1979). In longitudinal studies of infants, using parental reports, two orthogonal factors of language comprehension and production seemed to exist (Bates et al. 1988). Alcock and Krawczyk (2010) specifically investigated the link between oral motor control, and other motor and language abilities in 21-month-olds, and concluded that oral motor control was significantly associated with the grammatical complexity of utterances and with language production ability overall. They found no relationship between overall motor control and language comprehension ability at this age. In our study with school age children, oral motor control was linked to the production of novel words, but did not predict individual differences in the comprehension of syntactically complex sentences (Krishnan et al., in press). This evidence may appear contradictory, but taking a developmental perspective may provide some explanation. First, there may be specific motor behaviours that provide an opportunity to acquire and practise skills necessary for language. For example, rhythmic hand banging peaks around 28 weeks of age, and this is also when children start to produce reduplicated babbling (Iverson 2010). And, when rhythmic banging is delayed, as is the case in the neurodevelopmental disorder Williams syndrome, babbling and subsequent comprehension and production are also delayed (Masataka 2001). Understanding the specific skills that are likely to cause changes in behaviour during a particular time-window may therefore be necessary. P&G's model fails to provide an account of how comprehension/production processes for learning language might be integrated within specific skills over developmental time.

The second factor that must be considered in a model integrating comprehension/production processes is the nonmonotonicity of these developmental trajectories, which are consistent across children. For example, rhythmic banging is low in pre-babbling children, increases sharply as infants start to babble, and then declines as infants become experienced at babbling (Iverson 2010). Similar nonmonotonic trajectories are seen across other speech motor skills, for instance, in the variability of lip and jaw movements (Smith & Zelaznik 2004) or for the coordination of upper and lower lip movements (Green et al. 2000). The combination of gestures and language during development may have a similar nonmonotonicity, as event-related potential (ERP) evidence suggests children infants younger than 20 months interpret symbolic gestures and words similarly, but that gestures and words take on divergent communicative roles when infants are 26 months old (Sheehan et al. 2007). Although P&G suggest that experience may be important for learning inverse-forward model pairing, their model lacks explanations of how these kinds of trajectories might arise, how trajectories change with time, and what kind of input may be necessary for change. For example, these trajectories may arise due to some combination of the changes in contextual support that are needed while a skill is learnt, or the neural changes that occur during development.

Therefore, in the model that P&G outline, I agree that integrating knowledge about the production processes may help us understand more about language comprehension, but this would be possible only if the specificity of production-comprehension links and the developmental timing of their occurrence are taken into account.

## Integrate, yes, but *what* and *how*? A computational approach of sensorimotor fusion in speech

doi:10.1017/S0140525X12002634

Raphaël Laurent,<sup>a,b</sup> Clément Moulin-Frier,<sup>a,c</sup>  
Pierre Bessière,<sup>b,d</sup> Jean-Luc Schwartz,<sup>a</sup> and Julien Diard<sup>e</sup>

<sup>a</sup>GIPSA-Lab – CNRS UMR 5216, Grenoble University, 38402 Saint Martin D'Hères Cedex, France; <sup>b</sup>e-Motion team - INRIA Rhône-Alpes, 38334 Saint Ismier Cedex, France; <sup>c</sup>FLOWERS team - INRIA Bordeaux Sud-Ouest, 33405 Talence Cedex, France; <sup>d</sup>Laboratoire de Physiologie de la Perception et de l'Action – CNRS UMR 7152, Collège de France, 75005 Paris, France; <sup>e</sup>Laboratoire de Psychologie et NeuroCognition – CNRS UMR 5105, Grenoble University, 38040 Grenoble Cedex 9, France.

Raphael.Laurent@gipsa-lab.grenoble-inp.fr

clement.moulin-frier@inria.fr Pierre.Bessiere@College-de-France.fr

Jean-Luc.Schwartz@gipsa-lab.grenoble-inp.fr

Julien.Diard@upmf-grenoble.fr

[http://www.gipsa-lab.grenoble-inp.fr/page\\_pro.php?vid=1238](http://www.gipsa-lab.grenoble-inp.fr/page_pro.php?vid=1238)

[http://www.gipsa-lab.grenoble-inp.fr/~clement.moulin-frier/cv\\_en.html](http://www.gipsa-lab.grenoble-inp.fr/~clement.moulin-frier/cv_en.html)

<http://www.Bayesian-Programming.org>

<http://www.gipsa-lab.grenoble-inp.fr/~jean-luc.schwartz/>

<http://diard.wordpress.com/>

**Abstract:** We consider a computational model comparing the possible roles of “association” and “simulation” in phonetic decoding, demonstrating that these two routes can contain similar information in some “perfect” communication situations and highlighting situations where their decoding performance differs. We conclude that optimal decoding should involve some sort of fusion of association and simulation in the human brain.

In their target article, Pickering & Garrod (P&G) propose an ambitious model of language perception and production. It is centered on three main ingredients. First, it considers the complete hierarchy of layers of language processing, from message to semantics to syntax to phonology and finally, to speech. Second, it features predictive forward models, so that temporally extended sequences, such as whole sentences and dialogues, can be processed. Third, it features dual processing routes, the “association” route and “simulation” route, so that auditory and motor knowledge can be involved simultaneously, rejecting the classic dichotomy between perception and action processes.

In this commentary, we set aside the temporal and hierarchical aspects, and focus on the domain of speech perception and production, where sequences are typically short (e.g., syllable perception and production), and processing limited to phonological decoding. Even in this more restricted field, the age-old debate between purely motor-based accounts and purely sensory-based accounts of perception and production now appears to be a false dilemma (Schwartz et al. 2012). Indeed, neurophysiological and behavioral evidence strongly suggests a dual route account of information processing in the central nervous system, with both a direct, associative route and an indirect, simulation route. The target article amply documents the evidence, we do not repeat examples here.

In our view, the debate is now shifting toward the issue of the functional role of each route and their integration. That is to say, a central question of the debate asks *what* is integrated and *how* integration proceeds in the human brain.

We would argue that conceptual models such as proposed in the target article would unfortunately have a difficult time bringing light to these questions. To support this argument, we consider the question of perceptual decoding of phonetic units, for which we have developed a computational framework (Moulin-Frier et al. 2012) based on Bayesian programming (Bessière et al. 2008; Colas et al. 2010; Lebellet et al. 2004). With this framework, various models of speech perception can be simulated and quantitatively compared. One model is purely auditory, exploiting what P&G call “association.” A second model is purely motor, exploiting what they call “simulation.” A third one

is sensory-motor, integrating the association and simulation processes.

All of these models can then be implemented and compared in various experimental configurations. Three major results emerge from such comparisons.

1. Under some hypotheses, with perfectly identified communication noise and no difference between motor repertoires of the speaker and the listener (i.e., when conditions for speech communication are “perfect”), motor and auditory theories are indistinguishable. Therefore, the “association” and “simulation” routes provide exactly the same information in these perfect communication conditions. The reason is that, in our learning scenario, the auditory classifier is learned by association from data obtained through a motor production process, and possesses enough mathematical power of expression.

This casts an interesting light on the question of *what* information is encoded in the association and simulation routes: Labeling a box as an “association” route, in a conceptual model, is not enough to be certain that it is different, from an information processing point of view, from another box of the model. Computational descriptions however, by virtue of rigorous mathematical notation, have to be precisely defined, and their content can be systematically assessed. This also explains why behavioral evidence has historically not been able to discriminate between motor and auditory theories of perception and production: They are sometimes simply indistinguishable. Unfortunately, we believe this difficulty was not avoided in the target article, in particular when P&G detail experimental evidence for their model (e.g., target article, sect. 3.2.1, para. 7, “these four studies support forward modeling, but they do not discriminate between prediction-by-simulation and prediction-by-association”; and sect. 3.2.3, para. 6, “all of these findings provide support for the model of prediction-by-simulation [...]. Of course, comprehenders may also perform prediction-by-association [...].”).

2. In the general case where “perfect conditions” for communication are not met, mathematical comparison of the models emphasizes the respective roles of motor and auditory knowledge in various conditions of speech perception in adverse conditions. Therefore, the information provided by the “association” and “simulation” routes is more or less distinct and prominent depending on the communication conditions. In other words, this demonstrates that adverse conditions provide leverage for discriminating hypotheses about the perceptual and motor processes involved. This is convergent with recent findings from neuroimaging and transcranial magnetic stimulation (TMS) studies (D’Ausilio et al. 2012b; Meister et al. 2007; Zekveld et al. 2006), as well as computational studies (Castellini et al. 2011).

3. In any case, sensory-motor fusion provides better perceptual performance than pure auditory or motor processes. Therefore, complementarities of information provided by the “association” and “simulation” routes could be efficiently exploited in the framework of integrative theories such as those hinted at in the discussion of the target article. It is now obvious in the field of audiovisual perception that auditory and visual cues are complementary, with a great deal of work already done on sensor fusion. In our opinion, comparable work can now be done on *how* to integrate auditory and motor processes in speech perception. In this view, the proposal by P&G that “comprehenders emphasize whichever route is likely to be more accurate” (sect. 4, para. 6) can be regarded as a first candidate model, which would have to be made mathematically precise and compared with alternative explanations, possibly driven by neuroanatomical findings (e.g., both auditory and motor processes are performed automatically in parallel and compete, or they both bring information in an ongoing fusion process, etc.).

An obvious challenge, of course, is to bridge the gap between computational approaches such as ours, which are usually restricted to isolated syllable production and perception, and conceptual models as proposed in the target article, that tackle

continuous flows of speech and consider semantic, syntactic and phonology layers of processing.

However, in our view, the main challenge for future studies is first to assess *what* kind of information is present in “association” and “simulation” routes, and second, to better understand *how* computational fusion models, describing the integration of these two routes, can account for experimental neurocognitive data.

## Towards a complete multiple-mechanism account of predictive language processing

doi:10.1017/S0140525X12002646

Nivedita Mani<sup>a</sup> and Falk Huetting<sup>b,c</sup>

<sup>a</sup>Language Acquisition Junior Research Group, Georg-August-Universität Göttingen, 37073 Göttingen, Germany; <sup>b</sup>Max Planck Institute for Psycholinguistics, 6500 AH Nijmegen, The Netherlands; <sup>c</sup>Donders Institute for Brain, Cognition, and Behaviour, Radboud University Nijmegen, Nijmegen, The Netherlands.

nmani@gwdg.de falk.huetting@mpi.nl

**Abstract:** Although we agree with Pickering & Garrod (P&G) that prediction-by-simulation and prediction-by-association are important mechanisms of anticipatory language processing, this commentary suggests that they: (1) overlook other potential mechanisms that might underlie prediction in language processing, (2) overestimate the importance of prediction-by-association in early childhood, and (3) underestimate the complexity and significance of several factors that might mediate prediction during language processing.

Pickering & Garrod (P&G) propose a model of language processing that blends production and comprehension mechanisms in such a way as to allow language users to covertly predict upcoming linguistic input. They ascribe a central role to our production system covertly anticipating what the other person (or oneself) might be likely to say in a particular situation (prediction-by-simulation). A second prediction mechanism, they argue, is based on the probability of a word being uttered (given other input) in our experience of others’ speech (prediction-by-association). We agree with P&G that prediction-by-simulation and prediction-by-association are important mechanisms of anticipatory language processing but believe that they overlook other (additional) potential mechanisms, overestimate the importance of prediction-by-association in early childhood, and underestimate the complexity and significance of several mediating factors.

**Multiple mechanisms.** If we consider predictive language processing to be any pre-activation of the representational content of upcoming words, then only multiple-mechanism accounts of prediction that are even more comprehensive than P&G’s can account for the multifarious phenomena documented in the language processing literature to date. Schwanenflugel and Shoben (1985), for example, have argued that more featural restrictions of upcoming words are generated in advance of the input in high as opposed to low predictive contexts. These featural (e.g., semantic, syntactic) restrictions may then constrain what words are likely to come up. In other words, pre-activation of representational content is constrained by the featural restrictions generated in particular contexts (e.g., via spreading activation in a semantic network). Whereas such online generation of featural restrictions may be compatible with prediction-by-simulation (as P&G appear to point out), it is conceivable that such pre-activation can also happen independently of prediction-by-simulation and thus constitutes a third anticipatory mechanism.

Predictive language processing may also make use of certain heuristics and biases (cf. Tversky & Kahneman 1973). Borrowing the availability heuristic from decision-making research, upcoming words may be pre-activated simply because of an availability bias (certain words/representations may be very frequent or have occurred very recently). The analogy to decision-making

research is not as far-fetched as one may think: Tversky and Kahneman introduced a simulation heuristic in the 1970s according to which people predict the likelihood of an upcoming event by how easy it is to simulate it. Other heuristics may well be worth exploring (in line with the affect heuristic, emotionally charged words, e.g., stupid, boring, may be predicted more rapidly). Our point is simply that predictive language processing is likely to be complex and may make use of a set of rather diverse mechanisms (with many yet unexplored).

**Importance of prediction-by-simulation in development.** P&G suggest that analysis of children's prediction abilities might throw light on the distinction between prediction-by-association and prediction-by-simulation and place stronger emphasis on prediction-by-association in young children: Prediction-by-association might play a more important role when listeners and speakers have little in common with each other, such as the case of children listening to adults' talking.

In a recent experiment examining 2-year-olds' prediction abilities, however, we found that, consistent with prediction-by-simulation, only toddlers in possession of a large *production* vocabulary are able to predict upcoming linguistic input in another speaker's utterance (Mani & Huettig 2012; see Melzer et al. 2012, for similar results in action perception). Furthermore, if, as P&G suggest, covert imitation is the driving force of prediction-by-simulation, then 18-month-olds are equipped with the cognitive pre-requisites for covert imitation: Covert imitation can modulate infants' eye gaze behaviour around a (linguistically relevant) visual scene (Mani & Plunkett 2010; Mani et al. 2012) similar to adults' behaviour (Huettig & McQueen 2007). Prediction-by-simulation may also be an important developmental mechanism to train the production system (Chang et al. 2006). In sum, prediction-by-simulation appears to be crucial even early in development, and hence prediction-by-association is not necessarily the simple prediction mechanism which dominates early childhood.

**Mediating factors.** Finally, there are many mediating factors (e.g., literacy, working memory capacity, cross-linguistic differences) involved in predictive language processing whose interaction with anticipatory mechanisms have been little explored and whose importance, we believe, has been vastly underestimated. Mishra et al. (2012), for instance, observed that Indian high literates, but not low literates, showed language-mediated anticipatory eye movements to concurrent target objects in a visual scene. Why literacy modulates anticipatory eye gaze remains to be resolved, though literacy-related differences in associations (including low-level word-to-word contingency statistics, McDonald & Shillcock, 2003), online generation of featural restrictions, and general processing speed are likely to be involved. Similarly, Federmeier et al. (2002) found that older adults are less likely to show prediction-related benefits during sentence processing with a strong suggestion that differences in working memory capacity underlie differences in predictive processing. The influence of such mediating factors may greatly depend on the situation language users find themselves in: Anticipatory eye gaze in the visual world, for instance, requires the building of online models allowing for visual objects to be linked to unfolding linguistic information, places, times, and each other. Working memory capacity may be particularly important for anticipatory processing during such language-vision interactions (Huettig & Janse 2012).

More work is also required with regard to the specific representations which are pre-activated in particular situations. Event-related potential studies have shown that even the grammatical gender (van Berkum et al. 2005), phonological form (DeLong et al. 2005), and visual form of the referents (Rommers et al. 2013) of upcoming words can be anticipated. Most of these studies, however, have used highly predictive "lead-in" sentences. It also remains to be seen to which extent these specific representations are activated in weakly and moderately predictive contexts. Last but not least, languages differ dramatically in all levels of linguistic organisation (Evans & Levinson 2009). These cross-linguistic

differences are bound to have substantial impacts on the specifics (and degree) of anticipatory processing a particular language affords.

Future work could usefully explore the cognitive reality and relative importance of the potential mechanisms and mediating factors mentioned here. Even though Occam's razor may favour single-mechanism accounts, we conjecture that multiple-mechanism accounts are required to provide a complete picture of anticipatory language processing.

## Toward a unified account of comprehension and production in language development

doi:10.1017/S0140525X12002658

Stewart M. McCauley and Morten H. Christiansen

Department of Psychology, Cornell University, Ithaca, NY 14853.

[smm424@cornell.edu](mailto:smm424@cornell.edu) [christiansen@cornell.edu](mailto:christiansen@cornell.edu)

<http://cnl.psych.cornell.edu>

**Abstract:** Although Pickering & Garrod (P&G) argue convincingly for a unified system for language comprehension and production, they fail to explain how such a system might develop. Using a recent computational model of language acquisition as an example, we sketch a developmental perspective on the integration of comprehension and production. We conclude that only through development can we fully understand the intertwined nature of comprehension and production in adult processing.

Much like current approaches to language processing, contemporary accounts of language acquisition typically assume a sharp distinction between comprehension and production. This assumption is driven, in large part, by evidence for a number of asymmetries between comprehension and production in development. Comprehension is usually taken to precede production (e.g., Fraser et al. 1963), although there are certain instances in which children exhibit adult-like production of sentence types that they do not appear to comprehend correctly (cf. Grimm et al. 2011). Evidence for such asymmetries strongly constrains theories of language acquisition, challenging integrated accounts of development and, by extension, integrated accounts of adult processing. Hence, it is key to determine the plausibility of a unified framework for acquisition that is compatible with evidence for comprehension/production asymmetries.

Although Pickering & Garrod's (P&G's) target article may be construed as a useful point of departure in this respect, P&G pay scant attention to how such a unified system for comprehension and production might develop. As a result, they implicitly subscribe to a different, questionable distinction often made in the language literature: the separation of acquisition from adult processing. In light of this, and given the tendency of developmental psycholinguists to view comprehension and production as separate systems, we briefly sketch a unified developmental framework for understanding comprehension and production as a single system, instantiated by a recent usage-based computational model of acquisition (McCauley & Christiansen 2011; submitted). Importantly, our approach is consistent with evidence for comprehension/production asymmetries in development, even while uniting comprehension and production within a single framework.

Our computational model, like that of Chang et al. (2006), simulates both comprehension and production, but it goes beyond this and previous usage-based models (e.g., Borensztajn et al. 2009; Freudenthal et al. 2007) in that (a) it learns to do so incrementally using simple distributional information; (b) it offers broad, cross-linguistic coverage; and (c) it accommodates a range of developmental findings. The model learns from corpora of child and child-directed speech, acquiring item-based knowledge in a purely incremental fashion, through online learning using backward transitional probabilities (which infants track; cf. Pelucchi et al. 2009). The model uses peaks and dips in

transitional probabilities to chunk words together as they are encountered, incrementally building an item-based “shallow parse” as each incoming utterance unfolds. The model stores the word sequences it groups together, gradually building up an inventory of multiword chunks – a “chunkatory” – which underlies both comprehension and production. When the model encounters a multiword utterance produced by the target child of a corpus, it attempts to generate an identical utterance using only chunks and transitional probabilities learned up to that point. Crucially, the very same chunks and distributional information used during production are used to make predictions about upcoming material during comprehension. This type of prediction-by-association facilitates the model’s shallow processing of the input. The model’s comprehension abilities are scored against a state-of-the-art shallow parser, and its production abilities are scored against the target child’s original utterances (the model’s utterances must match the child’s).

The model makes close contact with P&G’s approach in that it uses information employed during production to make predictions about upcoming linguistic material during comprehension (consistent with recent evidence that children’s linguistic predictions are tied to production; cf. Mani & Huettig 2012). However, our approach extends P&G’s account from prediction to the acquisition and use of linguistic knowledge itself; comprehension and production rely upon a single set of statistics and representations, which are reinforced in an identical manner during both processes.

Moreover, our model’s design reflects recent psycholinguistic findings that have hitherto remained largely unconnected, but which, when viewed as complementary to one another, strongly support a unified framework for comprehension and production. First, the model is motivated by children’s use of multiword units in production (Bannard & Matthews 2008), which cautions against models of production in which words are selected independently of one another. The model’s primary reliance on the discovery and storage of useful multiword sequences follows this line of evidence. Second, the model is motivated by evidence that children, like adults, can rely on shallow processing and underspecified representations during comprehension (e.g., Gertner & Fisher 2012; Sanford & Sturt 2002). Shallow processing, supplemented by contextual information (e.g., tied to semantic and pragmatic knowledge) may often give children the appearance of comprehending grammatical constructions they have not yet mastered (and therefore cannot use effectively in production). The model exhibits this in its better comprehension performance; through chunking, the model can arrive at an item-based “shallow parse” of an utterance, which can then be used in conjunction with semantic and pragmatic information to arrive at a “good enough” interpretation of the utterance (Ferreira et al. 2002). On the production side, however, the model – like a child learning to speak – is faced with the task of retrieving and sequencing words and chunks in a particular order. Hence, asymmetries arise from differing task demands, despite the use of the very same statistics and linguistic units during both comprehension and production.

Such an abandonment of the “cognitive sandwich” approach to acquisition clearly has implications for adult processing. If, as we suggest and make explicit in our model, children learn to comprehend and produce speech by using the same distributional information and chunk-based linguistic units for both tasks, we would expect adults to continue to rely on a unified set of representations. This is corroborated by studies showing that, like children, adults not only rely on multiword units in production (Janssen & Barber 2012), but also use multiword sequences during comprehension (e.g., Arnon & Snider 2010; Reali & Christiansen 2007). This evidence further suggests that prediction-by-association may be more important for language processing than assumed by P&G, not just for children as indicated by our model, but also for adults. It is only by considering how the adult system emerges from the child’s attempts to comprehend and produce linguistic utterances that we can hope to reach a

complete understanding of the intertwined nature of language comprehension and production.

## What does it mean to predict one’s own utterances?

doi:10.1017/S0140525X12002786

Antje S. Meyer<sup>a,b</sup> and Peter Hagoort<sup>a,b</sup>

<sup>a</sup>Max Planck Institute for Psycholinguistics, 6500 AH Nijmegen, The Netherlands; <sup>b</sup>Radboud University Nijmegen, 6525 HP Nijmegen, The Netherlands.

antje.meyer@mpi.nl    peter.hagoort@mpi.nl  
www.mpi.nl

**Abstract:** Many authors have recently highlighted the importance of prediction for language comprehension. Pickering & Garrod (P&G) are the first to propose a central role for prediction in language production. This is an intriguing idea, but it is not clear what it means for speakers to predict their own utterances, and how prediction during production can be empirically distinguished from production proper.

Pickering & Garrod (P&G) offer an integrated framework of speech production and comprehension, highlighting the importance of predicting upcoming utterances. Given the growing evidence for commonalities between production and comprehension processes and for the importance of prediction in comprehension, we find their proposal timely and interesting.

Our comment focuses mainly on the role of prediction in language production. P&G propose that speakers predict aspects of their utterance plans and compare these predictions against the actual utterance plans. This monitoring process happens at each processing level, that is, minimally at the semantic, syntactic, and phonological level.

Given the important role of prediction in comprehension and the well-attested similarities between production and comprehension, the idea that prediction should play a role in speech production follows quite naturally. Nevertheless, to us the proposal that speakers predict their utterance plans does not have immediate appeal. This is because, in everyday parlance, prediction and the predicted event have some degree of independence. It is because of this independence that predictions may or may not be borne out. It makes sense to say a person predicts the outcomes of their hand or jaw movements, as these outcomes are not fully determined by the cognitive processes underlying the predictions, but depend, among other things, on properties of the physical environment that may not be known to the person planning the movement. Similarly, it makes sense to say that a listener predicts what a speaker will say because the speaker’s utterances are not caused by the same cognitive processes as those that lead to the listener’s prediction. Speaker and listener each have their own, private cognition and therefore the listener’s expectations about the speaker’s utterance may or may not be met.

We can predict our own utterances. For instance, based on memory of past experience, I can predict how I will greet my family. However, such predictions concern overt behavior rather than plans for behavior, and they occur offline rather than in parallel with the predicted behavior. Just like predictions about other persons, my predictions of my own utterances may or may not be borne out, depending on circumstances not known at the moment of prediction. I may, for instance, deviate from my predicted greeting if I find my family standing on their heads.

Such offline predictions of overt behavior differ from the predictions proposed by P&G. In their framework, speakers predict their utterance *plans* as they plan them, with prediction at each planning level running somewhat ahead of the actual planning. Importantly, the predictions are based on the same information as the predicted behavior, namely, the speaker’s intention

(target article, sect. 3.1, “production command” in Fig. 5) and involve closely related cognitive processes, although the plans are more detailed than the predictions and can therefore be created faster. With respect to the information encoded in both representations, plan and prediction will be identical. Discrepancies can only arise when the plan and/or prediction include a fault. This is different from predictions a listener might generate about a speaker’s upcoming utterance; no matter how well aligned speaker and listener are in a dialogue, their intentions are not identical, and their cognitive processes are not shared.

P&G invoke prediction during production to support self-monitoring. It is not entirely clear how the monitoring processes would work and how beneficial they would be. Key properties of the predicted representations are that they are more abstract and that they are created faster than the speech plans and not necessarily in the same order. This raises the issues of how it is decided which information to include in the prediction and what to omit and, given that planning and predicting need not follow the same time course, whether and how the cognitive processes leading to plans and predictions differ. It is also not clear why predictions would be more likely to be correct than plans, and how a speaker detects errors concerning features of the utterance that are not included in the predictions. Finally, as P&G point out, there is strong evidence for the involvement of forward modeling in motor planning, but there is as yet no empirical evidence demonstrating that this approach scales up in the way they envision. Finding this evidence is likely to be extremely challenging, as it will, for instance, involve separating the time course of planning and predicting and the properties of the planned and predicted representations.

The question of what and when to predict is also relevant for prediction during comprehension. P&G assume, correctly in our view, that in dialogue there is often not sufficient information in the mind of the comprehender to generate reliable predictions at all conceptual and linguistic levels. This results in two possibilities. One is that predictions are always made, but will often be highly unreliable, creating a need for correction that would not exist in the absence of prediction. The other option is that prediction occurs only if there is sufficient information for making a valid prediction. P&G seem to favor the latter option (“comprehenders make whatever linguistic predictions they can”; sect. 3.2, para. 1). However, their model should then specify a mechanism or procedure that determines when to predict and when not to do so. In passing, we note another gap in P&G’s proposal: According to P&G, predictions can be generated via an association route and a simulation route. But how are the contributions of the two prediction routes weighted and integrated? What happens if these predictions do not fully match?

In sum, we are not convinced that prediction plays a similar crucial role in speech production as it does in comprehension. Whereas my listener can at best guess (predict) what I might say next, as a speaker I know perfectly well where I am heading, and plans and predictions cannot be separated. Moreover, the account of prediction in both production and comprehension needs further specification of what triggers the decision to predict and of how predictions are derived.

## Is there any evidence for forward modeling in language production?

doi:10.1017/S0140525X1200266X

Myrto I. Mylopoulos<sup>a</sup> and David Pereplyotchik<sup>b</sup>

<sup>a</sup>Doctoral Program in Philosophy, City University of New York (CUNY), Graduate Center, New York, NY 10016; <sup>b</sup>Department of Philosophy, Hamilton College, Clinton, NY 13323.

myrto.mylopoulos@gmail.com dpereply@hamilton.edu  
http://www.myrtomylopoulos.com http://www.pereplyotchik.com/

**Abstract:** The neurocognitive evidence that Pickering & Garrod (P&G) cite in favor of positing forward models in speech production is not compelling. The data to which they appeal either cannot be explained by forward models, or can be explained by a more parsimonious model.

Pickering & Garrod (P&G) take production commands to be conceptual representations that encode high-level features – “information about communicative force (e.g., interrogative), pragmatic context, and a nonlinguistic situation model” (target article, sect. 3.1, para. 3). On their model, a production command is input directly to the production implementer, which outputs an utterance. But somewhere in between this input and output, there must be, in addition, an intermediate representation that specifies the low-level features of the utterance, for example, its phonological and phonetic features. In what follows, we will call this low-level production command the *utterance plan*. In the analogous motor control case, upon which P&G base their model, an utterance plan corresponds to a motor command, which specifies the low-level features of the bodily movement, and is output by the inverse model (Wolpert 1997).

We argue here that the evidence that P&G cite in favor of positing forward models in speech production is not compelling. More specifically, the data to which they appeal either cannot be explained by forward models, or can be explained by a more parsimonious model, on which the utterance plan and the sensory feedback are directly compared. On this alternative picture, there is no need to posit forward models.

P&G appeal to Heinks-Maldonado et al. (2006) to support their claim that forward modeling is used in speech production. They argue that the suppressed M100 signal in the condition where participants spoke and heard their own unaltered speech – compared with conditions in which their speech was distorted or they heard an artificial voice – is the result of the forward model prediction “canceling out” the matching auditory feedback from the utterance. They urge that “the rapidity of the effect suggests that speakers could not be comprehending what they heard and comparing this to their memory of their planned utterance” (sect. 3.1, para. 16). While this is indeed implausible, there is an alternative hypothesis that is not ruled out by the data: The attenuation effect results from a match between the utterance plan and auditory feedback. Such a comparison would take no more time than the purported comparison between the forward model prediction and auditory feedback. The same point applies to P&G’s discussion of the datum reported in Levelt (1983) concerning mid-word self-correction.

Some theorists (e.g., Prinz 2012, p. 238) have insisted that whatever states enter into the comparison with sensory feedback must have the same representational format as the feedback. Because an utterance plan must encode the low-level features of the utterance that it specifies, it arguably meets this criterion.

P&G also appeal to the results reported in Tourville et al. (2008), highlighting two features of that study. First, the compensation that participants make in response to distorted auditory feedback is rapid – “a hallmark of feedforward (predictive) monitoring (as correction following feedback would be too slow)” (sect. 3.1, para. 17). But rapid compensation can only be attributed to forward modeling when the forward model prediction is used in place of the auditory feedback during online control of behavior. The idea is that, by using the putative forward model prediction of the sensory feedback, the system need not wait for the auditory feedback. However, this cannot be the case in the experiment conducted by Tourville et al. (2008), because the distorted auditory feedback is externally induced at random, and therefore unpredictable. Participants must base their compensations on the distorted auditory feedback itself, since no prediction would be available in this type of case. Hence, however rapid their compensation, it cannot reflect the operation of forward modeling.

The second feature of the Tourville et al. (2008) study to which P&G appeal is that “the fMRI [functional magnetic resonance imaging] results identified a network of neurons coding

mismatches between expected and actual auditory signals” (sect. 3.1, para. 17). But while the fMRI results did identify a network of neurons that has been shown to be activated when auditory feedback from an utterance is distorted (Fu et al. 2006; Hashimoto & Sakai 2003; Hirano et al. 1997; McGuire et al. 1996), the further claim that these neurons code mismatches between forward model predictions (“expected” auditory signals) and actual auditory signals is unwarranted by the available neuroimaging data. All such data are equally consistent with the more parsimonious hypothesis that these neurons code mismatches between the utterance plan and the auditory feedback.

Finally, we are skeptical of P&G’s interpretation of the data in Tian and Poeppel (2010). The Tian and Poeppel study found activation in the auditory cortex in two conditions: after participants actually produced a syllable and after they merely imagined producing that same syllable. Following Tian and Poeppel, P&G interpret such activation as evidence of forward modeling. However, this activation may simply encode a general expectation that a sound will be heard, rather than specifically encoding the anticipated auditory feedback. Moreover, even if this activation were shown to be a representation of specific auditory feedback, it does not follow that the activation should be construed as a forward model prediction. It could instead subserve a mere *simulation* of the auditory feedback. In order to determine that this activation subserves a forward model *prediction*, as against a mere simulation, we would need evidence that it plays the relevant functional role—that is, it is both based on an efference copy *and* that it goes on to be compared to auditory feedback. Tian and Poeppel provide no evidence for the latter condition. Indeed, the relevant kind of forward model should not be operative in the passive listening condition of Tian and Poeppel’s study. So the fact that auditory cortex activations were found to be similar for listening passively to a sound and imagining producing that sound does not support the claim that the latter reflects forward modeling in particular. Finally, Tian and Poeppel’s framing of the issue is itself suspect. They cast forward model predictions as conscious personal-level states—mental images of a certain kind. One might reasonably doubt that such subpersonal states are ever present in the phenomenology that accompanies speech production.

## Inner speech as a forward model?

doi:10.1017/S0140525X12002798

Gary M. Oppenheim

Center for Research in Language, University of California San Diego,  
La Jolla, CA 92093-0526.

[goppenheim@crl.ucsd.edu](mailto:goppenheim@crl.ucsd.edu)

<http://crl.ucsd.edu/~goppenheim/>

**Abstract:** Pickering & Garrod (P&G) consider the possibility that inner speech might be a product of forward production models. Here I consider the idea of inner speech as a forward model in light of empirical work from the past few decades, concluding that, while forward models could contribute to it, inner speech nonetheless requires activity from the implementers.

Pickering & Garrod (P&G) argue that coarse predictions from forward models can help detect errors of overt speech production before they occur. This error-detecting function is often assigned to inner speech (e.g., Levelt 1983; Levelt et al. 1999; Nootboom 1969): the little voice in one’s head, better known for its role in conscious thought. It is therefore tempting to identify inner speech as a product of these forward models, with  $\hat{p} \rightarrow \hat{c}$  providing what we know as the internal loop. In fact, conceiving of inner speech as a forward model could elegantly address three key questions. First, why do we have inner speech at all? Inner speech is a

by-product of speakers’ need to control their overt verbal behavior. Second, why does inner speech develop so long after overt speech (e.g., Vygotsky 1962)? Inner speech develops as the speaker learns to simulate their verbal behavior, which may lag behind the ability to produce that behavior. And third, how are people able to produce inner speech without actually speaking aloud? If inner speech is simply the offline use of forward models ( $\hat{p} \rightarrow \hat{c}$ ), then speakers never need to engage the production and comprehension implementers ( $p \rightarrow c$ ) that are the traditional generators and perceivers of inner speech.

P&G’s framework would specifically address two more recently demonstrated qualities of inner speech. First, inner speech involves *attenuated* access to subphonemic representations. When people say tongue-twisters in their heads, their reported errors are less influenced by subphonemic similarities than their reported errors when saying them aloud (Oppenheim & Dell 2008; 2010; also Corley et al. 2011, as noted by Oppenheim 2012). For instance, /g/ shares more features with /k/ than with /v/, so someone trying to say GOAT aloud would more likely slip to COAT than VOTE, but this tendency is less pronounced for inner slips. As P&G note, this finding is predicted if the forward models underlying inner speech produce phonologically impoverished predictions (and thus might not reflect the production implementer). Second, inner speech is flexible enough to incorporate additional detail. Although inner slips show less pronounced similarity effects than overt speech, adding silent articulation is sufficient to boost their similarity effect, apparently coercing inner speech to include more subphonemic detail (Oppenheim & Dell 2010). Such flexibility could be problematic for models that assign inner speech to a specific level of the production process (e.g., Levelt et al. 1999), but P&G’s account specifically suggests that forward models simulate multiple levels of representation, so it might accommodate the subphonemic flexibility of inner speech by adding motoric predictions ( $\hat{p}[\text{sem}, \text{syn}, \text{phon}, \text{art}]$ ; forward models’ more traditional jurisdiction) that are tied to motor planning.

But forward model simulations cannot provide a complete account of inner speech. One would still need to use what P&G would call “the production implementer” (target article, sect. 3, para. 2). First, inner rehearsal facilitates overt speech production (MacKay 1981; Rauschecker et al. 2008; but cf. Dell & Repka 1992), suggesting that some aspects of the production implementer are also employed in inner speech. Second, there is abundant evidence that people easily detect their inner speech errors (Corley et al. 2011; Dell 1978; Dell & Repka 1992; Hockett 1967; Meringer & Meyer 1895, cited in MacKay 1992; Oppenheim & Dell 2008; 2010; Postma & Noordanus 1996). But since monitoring is described as the resolution of predicted and actual percepts (from forward models and implementers, respectively), it is unclear how one could detect and identify inner slips without having engaged the production implementer. (Conflict monitoring, e.g., Nozari et al. 2011, within forward models might at least allow error detection, but its use there seems to lack independent motivation, and still leaves the problem of how a speaker could identify the content of an inner slip.) Third, analogues of overt speech effects are often reported for experiments substituting inner-speech-based tasks. For instance, inner slips tend to create words, just like their overt counterparts (Corley et al. 2011; Oppenheim & Dell 2008; 2010), and their distributions resemble overt slips in other ways (Dell 1978; Postma & Noordanus 1996). And though inner and overt speech can diverge, they tend to elicit similar behavioral and neurophysiological effects in other domains (e.g., Kan & Thompson-Schill 2004), and their impairments are highly correlated (e.g., Geva et al. 2011). Though more ink is spilled cautioning differences between inner and overt speech, similarities between the two are the rule rather than the exception (at least for pre-articulatory aspects).

Given the impoverished character of P&G’s forward models, it seems difficult to account for such parallels without assuming a

role for production implementers in the creation of inner speech. Therefore, we could posit that inner speech works much like overt speech production, recalling P&G's acknowledgment that offline simulations could engage the implementers, actively truncating the process before articulation; forward models would supply a necessary monitoring component. This more-explicit account of inner speech allows us to question P&G's suggestion that the subphonemic attenuation of inner speech might reflect impoverishment of the forward model instead of the generation of an abstract phonological code by the production implementer. Having clarified the role of forward models as error detection, their suggestion now boils down to the idea that inner slips might be hard to "hear." Empirical work suggests that is not the case. Experiments using noise-masked overt speech (Corley et al. 2011) and silently mouthed speech (Oppenheim & Dell 2010) showed that each acts much like normal overt speech in terms of similarity effects (see also Oppenheim 2012). And, by explicitly modeling biased error detections, Oppenheim and Dell (2010) formally ruled out the suggestion that their evidence for abstraction merely reflected such biases. Thus, better specifying the role of forward models in inner speech allows the conclusion that the subphonemic attenuation of inner speech does have its basis in the production implementer. More generally, conceiving of forward models as *components* of inner speech can wed strengths of the forward model account with the fidelity of implementer-based simulations.

## Does what you hear predict what you will do and say?

doi:10.1017/S0140525X12002804

Mariella Pazzaglia

Dipartimento di Psicologia Sapienza University of Rome, 00185 Rome, Italy;  
IRCCS Fondazione Santa Lucia, 00179 Rome, Italy.

[mariella.pazzaglia@uniroma1.it](mailto:mariella.pazzaglia@uniroma1.it)

<http://mariellapazzaglia.com/>

<http://dippsy.psi.uniroma1.it/users/pazzaglia-mariella>

**Abstract:** I evaluate the bottlenecks involved in the simulation mechanism underpinning superior predictive abilities for upcoming actions. This perceptual-motor state is characterized by a complex interrelationship designed to make predictions using a highly fine-tuned and constrained motor operation. The extension of such mechanisms to language may occur only in sensorimotor circuits devoted to the action domain.

One of the most influential current theories suggests that higher-order socio-cognitive processes, such as mind/intention reading and the compound aspects of language, may be primarily "grounded" in sensorimotor brain (Barsalou 2008).

Inspired by multiple-duty cells originally discovered in the monkey (di Pellegrino et al. 1992), neuroimaging studies have proposed that the human brain is equipped with specific, rapid, and automatic mechanisms that share action execution and perception in a common representational domain (Aglioti & Pazzaglia 2010; 2011; Van Overwalle & Baetens 2009). According to Pickering & Garrod (P&G), this inherent bidirectional, functional, and anatomical link seems to monitor the perception of other agents' actions through predictive mechanisms. The authors suggest that a simulative process might also run an internal generative representation that serves predictions in response to linguistic input. Despite nearly two decades of intensive research on the inextricable link between the perception and execution of action, there are two key problems with action prediction through simulation and, consequently, with its application to language processing.

The first bottleneck concerns neurophysiological and cognitive constraints, as revealed by action predictive coding. Inferring the

intention of an action from a perceptual-motor code should imply accurate, one-to-one perceptual motor mapping between the goal and its respective kinematics (Kilner 2011). This is evident, for example, with a specific set of "action-constrained" single-cell recordings in monkeys, which fired during grasping for eating but not during grasping for placing (Fogassi et al. 2005). Upon exploring the various ways in which humans can reach and grasp, it appears that the kinematics precisely differ with respect to compatibility, or incompatibility, with the goal (e.g., drinking vs. passing) (Tretriluxana et al. 2008). Crucially, activity in the inferior frontal cortex of onlooking human individuals is modulated differentially when a model exhibits different intentions associated with the grasping action (Iacoboni et al. 2005). In order to achieve this overall intention, an individual selects the most appropriate movement that is compatible with the purpose of the action. Within this framework, it is clear that the motor representations are comparatively stable and can be arranged in a limited, pre-wired motor chain that is functionally interpreted in terms of motor intention (Rizzolatti & Sinigaglia 2007).

Speech sound representations, however, are highly variable; the linguistic message can be achieved with many speech sounds and, more questionably, the same speech units vary with their position within a word (Mottonen & Watkins 2009). In language processing, prediction by a simulation mechanism is plausible for articulatory representations consisting of a limited set of uniform elements that mainly differ in their serial positions and require precise selection and timing, and are at least pre-wired in two units (noun and verb) to form a complete sentence. This type of mechanism was reported for the hearing or reading of motor-related words/sentences, for which a growing number of studies have proposed a constant matching of input-output processes, however action-system mediated (Buccino et al. 2005; Pulvermüller & Fadiga 2010; Tettamanti et al. 2005). Moreover, several studies have documented how a rapid simulation process supports motor-related speech/language in the frontal motor cortex, ranging from the spontaneous imagery mechanism of tracking articulatory gestures to the complex motor aspects of action verbs or tool words that grant them their meaning (Pulvermüller 2005). Hence, the motor counterpart enables the matching of production and comprehension, extending motor-related sound identification to language (the "what" of speech recognition), which in turn leads to predictive coding. Given the role of motor system, it remains unclear whether language perception occurs in a more general cognitive-motor domain or is an independent representation interacting with the action system (Fadiga & Craighero 2006). In addition, an intense debate exists that intimately links language and action at the ontogenetic (Bates & Dick 2002) and phylogenetic (Toni et al. 2008; Zlatev 2008) levels. From an evolutionary perspective, studies have postulated that language initially evolved from manual gestures in the form of a system of manual skills, pantomime, and protosigns (Arbib 2005; Leroi-Gourhan 1964; 1965). The subsequent conventionalization of signs and the shift to vocal emblems has enabled the transition to more symbolic, alternative, and open systems of communication (Corballis 2009).

The second bottleneck refers to the controversial neural and functional evidence reported by neuropsychological analyses. In brain-damaged patients with significant deficits of execution, the ability to predict and understand an observed/heard action may be spared. Although recent studies indicated a positive correlation between deficits in perceived and performed actions (Buxbaum et al. 2005; Nelissen et al. 2010), many studies fail to provide straightforward evidence for direct matching between observed and executed actions (Hickok 2009a). Precise perceptual-motor coding, on which predictions must be planned, would explain the input-output associations of the impairment, but not the range of dissociations reported at both the group (Cubelli et al. 2000; Halsband et al. 2001; Negri et al. 2007; Pazzaglia et al. 2008a) or single-case (Pazzaglia et al. 2008b; Rumiati et al. 2001) level. Moreover, the published studies do not clarify

whether or not neurologic patients are still able to infer the intention of an observed action (Fontana et al. 2012), despite the disruption in the ability to mentally simulate movements (Pazzaglia et al. 2008a).

Although the dissociations between the neural and functional aspects of matching among input–output still need to be clarified in aphasic and apraxic patients (Pulvermüller et al. 2005; Pazzaglia 2013), the plausible implications of anatomical and clinical divergences cannot be ignored. Dissociation, rather than the association of neuropsychological deficits in brain-damaged patients, continues to be a highly sensitive verification technique that is necessary to exclude vitia and define the reliability boundaries of empirically viable theories.

Therefore, the range of possible dissociations between production and comprehension, which can occur in both action and language, is rather multifarious. In this conception, such dissociations are reliant upon higher-order sensorimotor experiences manifesting in the computational brain, namely: the intention to act; stable memory traces for different types of percepts; and the ecological and cultural conditions in which gestural, linguistic, and affective communication are implemented. Such processes could probably also interact with unique, more basic, low-level motor-resonance mechanisms (Mahon & Caramazza 2008). In particular, this can include the automatic selection of symbolization on which judgments regarding communication and predictions of appropriateness are based.

P&G discuss and emphasize studies that interweave the production and recognition of actions. However, they too quickly exclude the limits of prediction via the simulation of action. By not looking closely at the crucial roles of the physiological process (whereby predictions emerge through extremely fine-grained cognitive-motor operations) and the neurologic population (behavioral and anatomical disease is fully dissociable), the extension of such mechanisms to language may become unwarranted in situations where language does not call on cognitive-motor representations. A fruitful direction for tracking a complete theory of language processing must not only recognize the degree to which the processes underlying language and action are similar but should also discuss the intertwined and integrated aspects of this relationship, at least in a conceptual sense.

#### ACKNOWLEDGMENT

Mariella Pazzaglia is supported by the International Foundation for Research in Paraplegie (IRP, P133).

## Intentional strategies that make co-actors more predictable: The case of signaling

doi:10.1017/S0140525X12002816

Giovanni Pezzulo<sup>a,b</sup> and Haris Dindo<sup>c</sup>

<sup>a</sup>Istituto di Linguistica Computazionale “Antonio Zampolli,” CNR, 56124 Pisa, Italy; <sup>b</sup>Istituto di Scienze e Tecnologie della Cognizione, CNR, 00185 Roma, Italy; <sup>c</sup>Computer Science Engineering, University of Palermo, 90128 Palermo, Italy.

[giovanni.pezzulo@istc.cnr.it](mailto:giovanni.pezzulo@istc.cnr.it) [haris.dindo@unipa.it](mailto:haris.dindo@unipa.it)  
<http://www.istc.cnr.it/people/giovanni-pezzulo>  
<http://roboticslab.dinfo.unipa.it/index.php/People/HarisDindo>

**Abstract:** Pickering & Garrod (P&G) explain dialogue dynamics in terms of forward modeling and prediction-by-simulation mechanisms. Their theory dissolves a strict segregation between production and comprehension processes, and it links dialogue to action-based theories of joint action. We propose that the theory can also incorporate intentional strategies that increase communicative success: for example, signaling strategies that help remaining predictable and forming common ground.

We highly appreciate Pickering & Garrod’s (P&G’s) theory for four main reasons. First, P&G address dialogue from a joint

action perspective, rather than in isolation from action, perception, and interaction dynamics, as most linguistic theories do. P&G’s theory thereby points toward the naturalization of linguistic communication and might help our understanding of how it develops on top of the nonlinguistic “interaction engine” of our earlier evolutionary ancestors (Levinson 2006; Pezzulo 2011b).

Second, to explain how interlocutors predict one another and monitor the ongoing interaction, P&G use the notions of forward modeling and prediction-by-simulation (Dindo et al. 2011; Grush 2004; Wolpert et al. 2003). These notions are increasingly adopted in cognitive and social neuroscience; language studies could greatly benefit from linking to the same mechanistic framework. Note that P&G’s theory does not overlook the specificities of language processing, and it assumes that such a processing is structured along multiple levels: semantic, syntactic, and phonological.

Third, P&G provide a theoretically sound motivation for the use of production processes in (language) comprehension and comprehension processes in (language) production, dissolving a strict production-comprehension dichotomy. P&G’s analysis links well to a large body of evidence documenting the interactions between perception and production processes outside linguistic communication (Bargh & Chartrand 1999; Frith & Frith 2008; Sebanz et al. 2006a), making it an excellent entry point to study dialogue within an action-based framework.

Fourth, P&G’s theory explains how prediction and covert imitation increase communication success and produce the automatic alignment of linguistic representations, which they have extensively documented empirically (Garrod & Pickering 2004; Pickering & Garrod 2004).

These four reasons notwithstanding, however, we propose not only that producers and comprehenders predict and covertly imitate their interlocutors, but also that they adopt intentional strategies that make their actions and intentions more predictable and comprehensible (D’Ausilio et al. 2012a; Pezzulo 2011c; Sartori et al. 2009; Vesper et al. 2011). In other words, to increase communication success, we propose that they facilitate another’s predictive (but also perceptual, inferential, attention, and memory) processes. For example, they can adopt *signaling strategies* to remain predictable and form common ground.

There are various demonstrations in which producers modulate their behavior (e.g., loudness, choice of words, speech rate) – depending on contextual factors (e.g., amount of noise, prior knowledge or uncertainty of the comprehender) – to help the comprehender’s predicting and understanding. A well-known case is the exaggeration of the vowels in child-directed speech (“motherese,” Kuhl et al. 1997). Not only are these modulations used during teaching, but also when comprehension is difficult, as is evident to those finding themselves speaking more loudly and over-articulating in noisy environments (the so-called “Lombard effect”).

Signaling strategies can be characterized in terms of efficient management of resources (e.g., articulatory effort, time) within an optimization framework that optimizes the joint goal of communication success (Pezzulo 2011c; Pezzulo & Dindo 2011); see also Moore (2007). Signaling consists of the intentional modulation of one’s own behavior (e.g., over-articulation) so that, in addition to its usual pragmatic or communicative goals (e.g., informing the interlocutor), the performed action fulfills the additional goal of facilitating the interlocutor’s prediction and monitoring processes (e.g., lowers the uncertainty or cognitive load). Compared to an optimal action, this modulation comes at a “cost” (e.g., a motor cost associated with over-articulation). However, as signaling ultimately helps to maximize the joint goal of communicative success, it is part of a joint action optimization process and is not (necessarily) altruistic.

Within the optimization framework, a cost-benefit analysis determines the decision to signal or not. This implies that signaling should be more frequent in uncertain contexts, when prediction is harder. To assess the theory, we designed a (nonlinguistic) joint task in which signaling determined a motor cost. We reported that the producers’ signaling probability depended on

the comprehenders' uncertainty; producers stopped signaling when it was low (Pezzulo & Dindo 2011). As this was the case even when producers received no online feedback from the comprehender, we hypothesized that they maintained an internal model of the comprehender's uncertainty. Computational and empirical arguments suggest that the choice of signaling was strategic and intentional (although not necessarily conscious) rather than a by-product of interaction dynamics.

In repeated interactions, signaling and other forms of sensorimotor communication help in sharing representations and maintaining a reliable *common ground*, too (Clark 1996; Pezzulo & Dindo 2011; Sebanz et al. 2006a). For example, by emphasizing the change of topic during a dialogue, a producer can reduce the comprehender's uncertainty at the level of task representations (say, dialogue topics) rather than only relative to the current utterance and so form a common ground (i.e., "align" the task representations of the interlocutors). Considerations of parsimony apply, also: Although costly to maintain, the common ground facilitates the continuation of the interaction, as both interlocutors can use it to predict what comes next. Furthermore, alignment entails parsimony: The shared part need not be maintained in two distinct forward models (one for each interlocutor), but the same forward model can be used as a "production model" for one interlocutor and "recognition model" for the other. We modeled the interplay between shared task representations and online action predictions using a hierarchical generative (Bayesian) architecture in which the former provide *priors* to the latter, and signaling strategies can be used to share task representations intentionally (Pezzulo 2011c; Pezzulo & Dindo 2011).

Our proposals on signaling and joint action optimization can be expressed in the neurocomputational architecture of P&G's theory. For example, producers can modulate their behavior online by using prior knowledge and a forward model of the comprehender's comprehension processes (e.g., they can over-articulate if the environment is noisy or if they foresee prediction errors). In turn, comprehenders can use the feedback channel strategically to expose their mental states and uncertainty by using a forward model of the producer's prediction and monitoring process. Furthermore, producers can use offline predictions (briefly mentioned in P&G's theory) to maintain a model of the comprehender's prior knowledge and uncertainty, and to foresee the long-term communicative effects of their intended messages (a form of *recipient design*).

By incorporating these (and similar) mechanisms, P&G's theory can explain intentional strategies such as signaling that—we argue—act in concert with automatic processes of alignment and mutual imitation to facilitate prediction, align representations, and form common ground.

## Prediction is no panacea: The key to language is in the unexpected

doi:10.1017/S0140525X12002671

Hugh Rabagliati<sup>a</sup> and Douglas K. Bemis<sup>b</sup>

<sup>a</sup>Department of Psychology, University of Edinburgh, Edinburgh EH8 9JZ, Scotland, United Kingdom; <sup>b</sup>INSERM-CEA Cognitive Neuroimaging Unit, CEA/SAC/DSV/DRM/Neurospin Center, F-91191 Gif-sur-Yvette Cedex, France.

[hugh.rabagliati@ed.ac.uk](mailto:hugh.rabagliati@ed.ac.uk)    [Douglas.BEMIS@cea.fr](mailto:Douglas.BEMIS@cea.fr)

<https://sites.google.com/site/hughrabagliati/>

<http://www.unicog.org/people/doug.bemis/>

**Abstract:** For action systems, the critical task is to predict what will happen next. In language, however, the critical task is not to predict the next auditory event but to extract meaning. Reducing language to an action system, and putting prediction at center, mistakenly marginalizes our core capacity to communicate the novel and unpredictable.

The fluency and rapidity with which we make ourselves understood, especially within the context of a dialogue, demands explanation: It is astounding that speakers alternate with essentially a 0-ms gap between turns (Sacks et al. 1974). In their target article, Pickering & Garrod (P&G) rise to this challenge and put forward an interesting and cogent framework that addresses this pace, built upon an intertwining of production and comprehension processes in the service of language as an action system. This intertwining is the headline of their proposal, but the real explanatory meat lies in how these processes are jointly used: the creation and checking of forward models, also known as *predictions*, about upcoming linguistic events. These predictions speed comprehension, speed production, and thereby contribute to "the remarkable fluency of dialogue" (target article, sect. 5, para.1).

We agree that many aspects of language use (especially within dialogue) rely heavily on prediction, and in particular rely heavily on predictions about *observable* aspects of language, for example, a speaker's stops and starts. We therefore understand why P&G might conclude that language is a form of action and action perception, and why they then afford a central position to forward models and their ability to predictively monitor and control actions. But although we certainly concur that prediction plays an important explanatory role for theories of language, we cannot help feeling that the emphasis given to action-based prediction in this model—and prediction in general throughout much of recent psycholinguistics (Altmann & Mirković 2009; DeLong et al. 2005; Dikker et al. 2009; Hale 2001; Levy 2008)—is overstated. The truly unique and indispensable power of language does not lie in its ability to quickly communicate the foreseeable, but rather the *unforeseeable*; to rapidly transfer information that is novel, surprising, and unpredictable. In P&G's example, *The day was breezy so the boy went outside to fly a kite*, the critical phenomenon to explain is not why the last word *kite* is processed faster and more efficiently than the first word *day*, but rather how the initial phrase, *The day was breezy*, is understood at all, given its completely unpredictable location halfway through a paper on psycholinguistic theory. Unfortunately, this phenomenon is left unexplained by the framework of P&G, as it is not directly related to prediction or action perception. No amount of forward modeling can produce the meaning of this initial phrase, as this meaning is not predictable from the preceding context in any substantive way.

By ignoring this crucial, and to our minds primary, function of language—extracting meaning from novel expressions—P&G do not allow their framework to get off the ground. As their examples testify, their model works well when predictions about time  $t+1$  are generated during the last stages of a sentence. We see little evidence, however, that their model can explain what happens when  $t=0$ , at the beginning of a sentence: Prediction relies on context, and within P&G's prediction-centric framework, there is no provision for the initial creation of a context.

Ultimately, we think that solving this problem requires P&G to drop, or at least substantially soften, their characterization of language comprehension as a form of action perception. Understanding linguistic expressions goes far beyond perceiving the actions by which they are delivered, and often, as in the case of reading, there are no actions to be perceived at all. Neurologically, this dissociation between perception and understanding is clearly demonstrated by transcortical sensory aphasia (Boatman et al. 2000; Lichtheim 1885), where patients can repeat words (i.e., use perception and production) without understanding them. Language, then, cannot simply be an action system but rather a system capable of productively transforming incoming perceptual elements into complex internal mental representations that convey meaning.

To their credit, P&G recognize this problem to a certain degree and include "well-defined levels of linguistic representation, such as semantics, syntax, and phonology" (target article, sect. 1.3, para. 9) in their proposed cognitive architecture. However, it is

unclear how these levels operate within an action/action perception system, as P&G do not specify whether their attempt to “reject the cognitive sandwich” (sect. 1.2, para. 3) entails collapsing action, perception, and cognition into one system (as Hurley [2008a] proposed), or just action and perception. Either way, linguistic representations are too marginalized within the model and require considerable elaboration to capture the rich communicative possibilities of human language. The insistence that language is only an action system leaves P&G with a model that, although possibly eliminating the “cognitive sandwich,” limits any explanation of the core function of language.

We believe that accounts of language must first and foremost explain the understanding of *novel* expressions. In other words, it is not the primary function of language to align turns in a dialogue by facilitating the comprehension of predictable words, but rather to enable a listener to understand the meaning of a speaker. Any model of language must conform to this prioritization and place understanding at the center, flanked by supporting processes such as prediction.

To be sure, the type of forward models proposed by P&G may still play an important role within such a framework as control systems. In the same way that forward models can help explain how a dancer completes a complex fouetté en tournant without tumbling over, they can help explain the surprisingly error-free execution of complex, rapid, interlaced dialogues. But just as we would not expect theories of motor *control* to explain acts of motor *creativity* (such as how a dancer improvises), we should not expect an analogous theory to explain the core creative aspects of language: the algorithms by which an entirely unexpected sentence can be integrated and understood, or by which a complex novel thought becomes articulated as a sentence.

In sum, we do not doubt that people make predictions during language use, quite possibly through the construction and evaluation of forward models. We just do not believe that these predictions comprise the foundation stones of a psychological theory of communication. Rather, we believe psycholinguists should focus on the representations these forward models are computed over, the representations that allow creative linguistic thought.

## Memory and cognitive control in an integrated theory of language processing

doi:10.1017/S0140525X12002683

L. Robert Slevc<sup>a</sup> and Jared M. Novick<sup>b</sup>

<sup>a</sup>Department of Psychology and <sup>b</sup>Center for Advanced Study of Language, University of Maryland, College Park, MD 20742.

slevc@umd.edu jnovick1@umd.edu

http://lmcl.umd.edu

**Abstract:** Pickering & Garrod’s (P&G’s) integrated model of production and comprehension includes no explicit role for nonlinguistic cognitive processes. Yet, how domain-general cognitive functions contribute to language processing has become clearer with well-specified theories and supporting data. We therefore believe that their account can benefit by incorporating functions like working memory and cognitive control into a unified model of language processing.

Pickering & Garrod (P&G) offer an integrated model of language processing that subsumes production and comprehension into a single cognitive framework, treating language as a form of action and action perception (cf. Clark 1996). This model draws on work linking prediction to language comprehension (e.g., Rhode et al. 2011) and production (e.g., Dell et al. 1997), and fits with the more general idea that we interpret our world not only by analyzing incoming information, but also by initiating proactive processes of prediction and expectation (Bar 2009).

Although memory processes are not explicit in P&G’s framework, the model invokes the maintenance and evaluation of multiple predictions and percepts, and relies on the retrieval of contextual information to create forward, anticipatory models of individuals’ linguistic and nonlinguistic actions. Memory and other cognitive functions are presumably an important part of these processes. A large body of work has investigated how language processing interfaces with other cognitive abilities but, like most psycholinguistic research, this has progressed mainly independently in studies of language comprehension and production. Despite this divide, recent work is converging on similar conclusions about the types of nonlinguistic cognitive systems that are critically involved in language production and comprehension. This suggests that the role of these cognitive systems might fruitfully be included in the forward modeling processes advocated in P&G’s framework. We highlight how a few aspects of this framework might draw on other cognitive systems.

Generating a prediction (of one’s own or another’s speech) relies heavily on memory processes. Indeed, anticipating how an utterance or a discourse will unfold necessarily depends on the rapid coordination of considerable linguistic and contextual evidence (Altmann & Kamide 1999; Tanenhaus 2007). To predict effectively (and hence avoid confusion or misinterpretation), current input must be linked to representations in working memory and in a longer-term store of prior experience. Moreover, individuals must be able to update and override these representations as new input is encountered moment by moment.

In the case of prediction-by-association, language users must retrieve situation-relevant information and schemas from memory as well as encode relevant information for use in future associative predictions. It is, therefore, unsurprising that the ease with which interlocutors can successfully encode and retrieve relevant associations in memory relates to how successfully they can align their discourse models, both in terms of the utterance choices that speakers make (Horton & Gerrig 2005) and the interpretations that listeners reach (Brown-Schmidt 2009).

Prediction-by-simulation, too, likely relies on memory processes. For example, the accessibility of information in memory influences how and when information is produced (Slevc 2011), and because prediction-by-simulation relies on internal production mechanisms, memory-based accessibility must also influence the prediction of others’ speech. This is indeed the case. For example, anaphor resolution is sensitive to the cognitive prominence of antecedents (Cowles et al. 2007), and more accessible syntactic structures are easier to parse (Branigan et al. 2005). Additionally, irrelevant information active in memory can interfere with both production (Slevc 2011) and parsing (Fedorenko et al. 2006), which in some cases could be construed as interference with one’s successful prediction of upcoming material in real time.

In a sense, memory underlies the generation of predictions – linguistic and otherwise – and, conversely, it is when predictions are not met that linguistic information is better learned or encoded into memory (e.g., Chang et al. 2006). There is, therefore, a tight linkage of memory and language processes; in fact, the processes of forward modeling involved in language processing may even be the foundation for much of our verbal memory ability (cf. Acheson & MacDonald 2009).

But it is not just the act of generating predictions that relies on nonlinguistic cognitive processes. Another crucial component of P&G’s model is *monitoring*, that is, comparing predicted to observed utterance precepts. This comparison presumably involves a process of detecting mismatch (or conflict) and resolving any discovered incompatibility. Mounting evidence suggests that conflict detection and its resolution via cognitive control plays an important role in both language comprehension and production (Novick et al. 2009). During comprehension, conflict is a natural by-product of incremental parsing: When late-arriving evidence is inconsistent with a reader’s or listener’s current representation of sentence meaning, conflict resolution and cognitive

control functions deploy to revise earlier processing commitments (Novick et al. 2005). Presumably, this applies to the monitoring function as well: Conflict resolution processes must adjudicate when an utterance precept is inconsistent with a speaker's or listener's expectation.

Linguistic conflict resolution functions depend on the involvement of the left inferior frontal gyrus (IFG), an area recruited when conflict must be resolved during nonlinguistic memory tasks (Jonides & Nee 2006). If conflict resolution underlies processing in a shared production/comprehension system, then deficits in these conflict resolution functions (e.g., in patients with circumscribed lesions to the left IFG) should yield both expressive and receptive language deficits when linguistic representations conflict. This is indeed the case: Such patients are known to have selective memory impairments when conflict/interference demands are high (Hamilton & Martin 2007; Thompson-Schill et al. 2002), and they also suffer concomitant production and comprehension impairments under similar conditions (Novick et al. 2009).

In sum, we believe that an important extension of P&G's model is to consider how language processing interfaces with other cognitive systems such as working memory and cognitive control. This raises a number of questions; for example, how general or specific are the cognitive systems involved in prediction and monitoring? If domain-general, which domain-general mechanisms are involved – for example, what are the roles of implicit and explicit memory, and do other executive functions contribute? Consideration of these types of issues is likely to lead toward a more fully integrated theory of language processing and of cognitive function more generally.

## The poor helping the rich: How can incomplete representations monitor complete ones?

doi:10.1017/S0140525X12002695

Kristof Strijkers,<sup>a</sup> Elin Runnqvist,<sup>b</sup> Albert Costa,<sup>c</sup> and Phillip Holcomb<sup>d</sup>

<sup>a</sup>Laboratoire de Psychologie Cognitive, CNRS and Université d'Aix-Marseille, 13331 Marseille, France; <sup>b</sup>Departamento de Psicología Básica, Universitat de Barcelona, Barcelona 08035, Spain; <sup>c</sup>Universitat Pompeu Fabra, Center for Brain and Cognition, ICREA, Barcelona 08018, Spain; <sup>d</sup>Department of Psychology, Tufts University, Medford, MA 02155.

kristof.strijkers@gmail.com elin\_runnqvist@yahoo.es  
costalbert@gmail.com pholcomb@tufts.edu

**Abstract:** Pickering & Garrod (P&G) propose that inner speech monitoring is subserved by predictions stemming from fast forward modeling. In this commentary, we question this alignment of language prediction with the inner speech monitor. We wonder how the speech monitor can function so efficiently if it is based on incomplete representations.

Pickering & Garrod's (P&G's) integrated account of language production and comprehension brings forward novel cognitive mechanisms for a range of language processing functions. Here we would like to focus on the theoretical development of the speech monitor in P&G's theory and the evidence cited in support of it. The authors propose that we construct forward models of predicted percepts during language production and that these predictions form the basis to internally monitor and if necessary correct our speech. This view of a speech monitor grounded in domain-general action control is refreshing in many ways. Nevertheless, in our opinion it raises a general theoretical concern, at least in the form in which it is implemented in P&G's model. Furthermore, we believe that it is important to emphasize that the evidence cited in support of the use of forward modeling in speech monitoring is suggestive, but far from directly supporting of the theory.

In general terms, we question the rationale behind the proposal that incomplete representations constitute the basis of speech monitoring. A crucial aspect of P&G's model refers to timing. Because forward representations are computed faster than the actual representations that will be used to produce speech, the former serve to correct potential deviations in the latter representations. To ensure that the forward representations are available earlier than the implemented representations, P&G propose that the percepts constructed by the forward model are impoverished, containing only part of the information necessary to produce speech. But how can speech monitoring be efficient if it relies on "poor" representations to monitor the "rich" representations? For instance, a predicted syntactic percept could include grammatical category, but lack number and gender information.

In this example, it is evident that if the slower production implementer is erroneously preparing a verb instead of a noun, the predicted representation coming from the forward model might indeed serve to detect and correct the error prior to speech proper. However, if the representation prepared by the production implementer contains a number or gender error, given that this information is not specified in the predicted percept (in this example), then how do we avoid these errors when speaking? If the predicted language percepts are assumed to always be incomplete in order to be available early in the process, it is truly remarkable that an average speaker produces only about 1 error every 1,000 words (e.g., Levelt et al. 1999). Therefore, although prediction likely plays an important role in facilitating the retrieval of relevant language representations (e.g. Federmeier 2007; Strijkers et al. 2011) and hence could also serve to aid the speech monitor, a proposal that identifies predictive processes with the inner speech monitor seems problematic or at least underspecified for now.

Besides the above theoretical concern regarding the use of incomplete representations as the basis of speech monitoring, also the strength of the evidence cited to support the use of forward modeling in speech production seems insufficient at present. The three studies discussed by P&G to illustrate the usage of efference copies during speech production (i.e., Heinks-Maldonado et al. 2006; Tian & Poeppel 2010; Tourville et al. 2008), demonstrate that shifting acoustic properties of linguistic elements in the auditory feedback given to a speaker produces early auditory response enhancements. Although these data are suggestive and merit further investigation, showing that refference cancellation generalizes to self-induced sounds does not prove that forward modeling is used for language production *per se*. It merely highlights that a mismatch between predicted and actual self-induced sounds (linguistic or not) produces an enhanced sensorial response just as in other domains of self-induced action (e.g., tickling). As for now, no study has explored whether auditory suppression related to self-induced sounds is also sensitive to purely linguistic phenomena (e.g., lexical frequency) or to variables known to affect speech monitoring (e.g., lexicality). This leaves open the possibility that the evidence cited only relates to general sensorimotor properties of speech (acoustics and articulation) rather than the monitoring of language proper.

In addition, the temporal arguments put forward by P&G to conclude that these data cannot be explained by comprehension-based accounts and instead support the notion of speech monitoring through prediction are premature. For instance, P&G take the speed with which self-induced sound auditory suppression occurs (around 100 ms after speech onset) as an indication that speakers could not be comprehending what they heard and argue that this favors a role of forward modeling in speech production. But, the speed with which lexical representations are activated in speech perception is still a debated issue and some studies provide evidence for access to words within 100 ms (e.g., MacGregor et al. 2012; Pulvermüller & Shtyrov 2006). In a similar vein, P&G rely on Indefrey and Levelt's (2004) temporal estimates of word production to argue in favor

of speech monitoring through prediction. However, this temporal map is still hypothetical, especially in terms of the latencies between the different representational formats (see Strijkers & Costa 2011). More generally, one may question why P&G choose to link the proposed model, intended to be highly dynamical (rejecting the “cognitive sandwich”), with temporal data embedded in fully serial models. Indeed, if one abandons the strictly sequential time course of such models and instead allows for fast, cascading activation of the different linguistic representations, not only do the arguments of P&G become problematic, but also the notion of a slow production/comprehension implementer being monitored by a fast and incomplete forward model loses a critical aspect of its theoretical motivation.

To sum up, we believe that theoretical development of the speech monitor in P&G’s integrated account of language production and comprehension faces a major challenge since it needs to explain how representations that lack certain dimensions of information can serve to detect and correct errors to such a high – almost errorless – degree. Furthermore, it is important to acknowledge that as it stands, the evidence used in support of this proposal could just as easily be reinterpreted in other terms, highlighting the need of direct empirical exploration of P&G’s proposal.

## When to simulate and when to associate? Accounting for inter-talker variability in the speech signal

doi:10.1017/S0140525X12002701

Alison M. Trude

Department of Psychology, University of Illinois at Urbana–Champaign,  
Champaign, IL 61820.

trude1@illinois.edu

**Abstract:** Pickering & Garrod’s (P&G’s) theory could be modified to describe how listeners rapidly incorporate context to generate predictions about speech despite inter-talker variability. However, in order to do so, the content of “impoverished” predicted percepts must be expanded to include phonetic information. Further, the way listeners identify and represent inter-talker differences and subsequently determine which prediction method to use would require further specification.

A hallmark of speech perception is that despite inter-talker variability on dimensions including rate, pitch, and phonetic variation due to accents, comprehension usually proceeds quickly and with little conscious effort. P&G’s theory provides a potential means of accommodating this variability by including context in the inverse model during comprehension; however, although the theory is potentially compatible with findings on talker adaptation, in order to generate testable hypotheses in this domain, it must more precisely specify what information is included in listeners’ predictions and how listeners assess talkers’ speech to determine which prediction route is more appropriate for the current input.

According to P&G’s model, listener-generated predictions are impoverished, which suggests that they do not include fine-grained phonetic detail. However, a large body of research shows that not only do listeners use fine-grained acoustic-phonetic details online while processing speech (McMurray et al. 2009; Salverda et al. 2003), but that this phonetic detail can also affect listeners’ subsequent productions (Nielsen 2011). If P&G wish to capture how listeners become phonetically aligned, allowing improved perception of an individual’s speech over time, the definition of “context” must be expanded to include phonetic details, as well as listeners’ previous experiences with a particular talker or group of talkers.

Another question raised by the model is how listeners’ use of the prediction-by-simulation and prediction-by-association routes

would vary as a consequence of talker identity. Consider, for example, an eye-tracking experiment from our laboratory showing rapid comprehension of a regional accent (Trude & Brown-Schmidt 2012). In this study, participants heard two talkers: a male with American English dialect in which /æ/ raises to /e/ only before /g/ (e.g., *tag* [teɪg], but *tack* [tæk]), and a female without this shift. On critical trials, participants viewed four images: a /k/-final target (e.g., *tack*), a /g/-final cohort competitor (e.g., *tag*), and two unrelated pictures. Participants clicked on the image named by one of the talkers. The results indicated that when listening to the male talker, participants fixated *tag* less upon hearing *tack*. Participants ruled the competitor out more quickly because the way that the male talker would have pronounced its vowel was not consistent with the input. Hence, we observed that listeners mentally represented an unfamiliar regional accent and used their knowledge rapidly enough to influence processing of a single word (see also Dahan et al. 2008).

P&G argue that listeners rely more on simulation when the talker is similar to them; however, the theory does not specify what degree of similarity is necessary for listeners to be able to use prediction-by-simulation during comprehension and when prediction-by-association is necessary. Additionally, for this model to generate testable hypotheses, it must specify how listeners assess the input in order to determine whether to use simulation or association, and the basis and frequency on which they update their assessments. According to P&G, context is determined using “information about differences between A’s speech system and B’s speech system” (target article, sect. 3.2, para. 3; Fig. 6 caption), suggesting that listeners engage in a comparison process. However, the details of that process, and how it aligns with current theories of speech perception, are unclear.

At a phonological level, it is possible that our participants would have been able to use the simulation route to predict the male’s vowel shift since, as native English speakers, the vowel /e/ is part of their own phonological system. It could also be the case, however, that our participants used the association route since their own phonological representation of *tag* includes an /æ/, rather than an /e/. At the same time, because our talkers and participants were all American English speakers, their speech was quite phonologically similar. Would this similarity have allowed our participants to use the simulation route most of the time, perhaps switching to association only for the critical vowel shift? Considering that the two talkers alternated randomly in our study, and that certain features of their speech may have been more or less like those of a given participant at different points in a single word, it seems as if it would have been necessary for the participant to constantly re-evaluate which prediction route was more suitable from moment to moment. This process would likely be too slow to implement and still produce the rapid online adaptation effects that we, and others, have observed.

A further question is whether listeners’ derived production commands are actually the same representations governing overt imitation in cases in which the talker’s and listener’s speech vary. In the model, the listener’s derived production command is generated after the percept has passed through the inverse model (which includes context), and therefore should include information about the talker’s voice; however, it appears that this command is also used to generate overt imitation. If so, it seems that listeners should be able to imitate whatever features of the talker’s speech they are able to predict (except when physiological differences prevent them from doing so). However, there are many cases in which a listener may *understand* a talker’s speech without readily producing certain features of it (Mitterer & Ernestus 2008). The use of impoverished representations during comprehension could explain listeners’ failure to overtly imitate these features; however, the fact that listeners can use fine-grained acoustic detail during comprehension seems at odds with this explanation. Furthermore, it has been shown that listeners *can* accommodate sub-phonemic features during imitation (Nielsen

2011), though they may do so to varying degrees (Babel 2012). Therefore, it would seem the theory needs a mechanism accounting for the dissociation between listeners' use of phonetic detail during comprehension and production.

In conclusion, P&G's theory can potentially explain talker-specific adaptation during comprehension because it allows a role for context while making predictions about a talker's speech. Furthermore, the rapid generation and implementation of representations is consistent with work using online methods that show talker-specific adaptation over the course of a single word. However, there are many open questions that remain about how listeners represent and predict the acoustic features of individuals' speech that must be addressed to make this a useful model of talker adaptation.

#### ACKNOWLEDGMENT

The writing of this commentary was supported by a National Science Foundation Graduate Research Fellowship, no. 2010084258.

## What is the context of prediction?

doi:10.1017/S0140525X12002713

Si On Yoon<sup>a</sup> and Sarah Brown-Schmidt<sup>b</sup>

<sup>a</sup>Department of Psychology, University of Illinois, Champaign, IL 61820;

<sup>b</sup>Department of Psychology, Beckman Institute for Advanced Science and Technology, University of Illinois, Champaign, IL 61820.

syoon@illinois.edu    brownsch@illinois.edu

**Abstract:** We agree with Pickering & Garrod's (P&G's) claim that theories of language processing must address the interconnection of language production and comprehension. However, we have two concerns: First, the central notion of context when predicting what another person will say is underspecified. Second, it is not clear that P&G's dual-mechanism model captures the data better than a single-mechanism model would.

We agree with Pickering & Garrod's (P&G's) claim that models of language use must take into account the fact that production and comprehension processes are interwoven in time and interconnected as in the case of split turns. Indeed, the most basic form of language use is arguably conversation, in which interlocutors act both as producers and addressees, and each type of act involves elements of the other.

The importance of examining dialogic processes has become apparent following a surge of interest in studying language in natural settings (Pickering & Garrod 2004; Trueswell & Tanenhaus 2005). This interest has generated new, significant findings in conversation including development of techniques for independently quantifying and predicting the degree of coordination in conversation (Richardson et al. 2007), as well as evidence that coordinated contextual representations facilitate use of potentially ambiguous referential expressions (Brown-Schmidt & Tanenhaus 2008). Similarly, experiments in noninteractive settings increasingly focus on dialogue-relevant questions such as how representations of others' mental states guide processing (Ferguson et al. 2010).

A central feature of language use is that it is produced and understood with respect to a particular context that constrains both what we say and how we say it. In conversation, the basic context is often assumed to be the interlocutors' common ground (Clark 1996) with conversational efficiency increasing as common ground grows (Wilkes-Gibbs & Clark 1992). According to P&G's proposal, context plays a key role in circumscribing prediction and, as a result, conversational efficiency.

According to P&G, listeners predict upcoming utterances using one of two mechanisms, *simulation* and *association*. Listeners use the simulation mechanism with familiar partners, and the association mechanism when they perceive themselves as being

different from their partner, as in adult-child conversations, and in reading, where the addressee does not typically speak. In what follows, we argue that with both familiar and unfamiliar partners, and in talking and reading alike, language users make sophisticated predictions about future language use, based on available contextual information.

P&G argue that an addressee using the simulation mechanism will predict an utterance using an inverse model and contextual information—a prediction about what the *speaker* would say. P&G do not specify the details of this contextual information. However, to predict what the speaker would say requires assessing the *speaker's* context, where context must be broadly defined to include both local perceptual information as well as historical information about the person's dialect and past experience. After all, depending on a person's age and regional dialect, a given semantic meaning might be expressed as *great*, *rad*, or *wicked*. Similarly, in cases where a potential alternative referent is occluded from the speaker's but not the addressee's view (e.g., the speaker sees one cup whereas the addressee sees two), the context would predict different referential expressions depending on which perspective was used. We read P&G as saying that in each of these cases, on the simulation mechanism, the addressee would predict what the speaker will say from the speaker's perspective, predicting "wicked" or "the cup," even though "wicked" might connote a negative valence to the addressee (rather than the intended positive valence), and even though "the cup" would be ambiguous from the addressee's perspective. Therefore, even though the prediction is executed using the addressee's production system, the entire prediction process would have to be tailored to the addressee's beliefs about the speaker's context.

In our view, the process would be no different on an association view in which addressees predict based on previous perceptual experience. P&G propose that association occurs when interlocutors are dissimilar or the production system is not engaged (e.g., in reading). However, even when unfamiliar interlocutors have different perspectives, overwhelming evidence now suggests that listeners do not progress egocentrically, but instead take into account information about their partner's context (e.g., Hanna et al. 2003; Heller et al. 2008). Similarly, as in live conversation, readers make rapid predictions when reading (Federmeier & Kutas 1999) and tailor referential interpretation based on representations of the number of entities in the discourse context (Greene et al. 1992; Nieuwland et al. 2007). Hence, it would seem that prediction-by-association would have to be tailored to the particular context of language use, just as in simulation. What advantage, then, is gained by positing a second mechanism to prediction? P&G suggest association might not afford rapid turn-taking, however this seems less of an argument to posit this mechanism than an argument against it, given that turn-taking is, in fact, rapid. P&G also suggest that individuals might choose to use either association, simulation, or a combination of the two; however, it is unclear how these decisions would be made, and how the outputs of these mechanisms would be integrated during real-time processing.

In conclusion, we applaud P&G's emphasis on the way production and comprehension are interwoven in natural communication. However, in emphasizing the remarkable skill needed to produce, for example, a split turn, the authors overlook potential redundancy in the dual-mechanism proposal. Indeed, the evidence suggests that in a large variety of circumstances, interlocutors integrate context and common ground into processing predictions. The accuracy, speed, and type of prediction seem to be determined largely by factors such as the quality of the listener's estimation of the speaker's context, and whether attending to one's partner's context is relevant to the communicative goals (Yoon et al. 2012). Hence, the determining factor in the quality of prediction should be seen as context-modeling, rather than a decision to use one mechanism in a hypothesized processing architecture. Understanding the mechanisms that determine the

context of conversation, and the degree to which the contexts of the speaker and listener are coordinated, then, would seem to be a central goal for understanding dialogic processes.

#### ACKNOWLEDGMENTS

Preparation of this article was partially supported by NSF grant no. BCS 10-19161 to S. Brown-Schmidt. Thank you to Jennifer Roche and Kara Federmeier for helpful discussions.

## Authors' Response

### Forward models and their implications for production, comprehension, and dialogue

doi:10.1017/S0140525X12003238

Martin J. Pickering<sup>a</sup> and Simon Garrod<sup>b</sup>

<sup>a</sup>Department of Psychology, University of Edinburgh, Edinburgh EH8 9JZ, United Kingdom; <sup>b</sup>University of Glasgow, Institute of Neuroscience and Psychology, Glasgow G12 8QT, United Kingdom.

[martin.pickering@ed.ac.uk](mailto:martin.pickering@ed.ac.uk)

[simon@psy.gla.ac.uk](mailto:simon@psy.gla.ac.uk)

<http://www.psy.ed.ac.uk/Staff/academics.html:PickeringMartin>

<http://staff.psy.gla.ac.uk/~simon/>

**Abstract:** Our target article proposed that language production and comprehension are interwoven, with speakers making predictions of their own utterances and comprehenders making predictions of other people's utterances at different linguistic levels. Here, we respond to comments about such issues as cognitive architecture and its neural basis, learning and development, monitoring, the nature of forward models, communicative intentions, and dialogue.

Our target article proposed a novel architecture for language processing. Rather than isolating production and comprehension from each other, we argued that they are closely linked. We first claimed that people predict their own and other people's actions. In a similar way, we argued that speakers predict their own utterances and comprehenders predict other people's utterances at a range of different linguistic levels.

The commentators made a wide range of perceptive points about our account, and we thank them for their input. We have divided their arguments into seven sections. We first respond to comments about the relation between production, comprehension, and other cognitive processes; in the second section, we turn to questions about the neural basis for our account. We said very little about learning and development in the target article, and our third section responds to those commentators who considered its implications for these issues. In the fourth section, we address the more technical issue of the nature of the representations created by forward modelling and how they are compared with implemented representations during monitoring. We respond, in the fifth section, to the commentators who remarked on the nature of prediction-by-simulation and its relationship to prediction-by-association. Finally, in sections six and seven we look at broader questions relating to the scope of the account: the nature of communicative intentions, and the implications of the account for dialogue.

### R1. Production, comprehension, and the "cognitive sandwich"

Our account has the overall goal of integrating production and comprehension. Our specific models in [Figures 5–7](#) (target article) are incompatible with traditional separation of production and comprehension as shown in [Figure R1](#) (repeated here from the target article). Some commentators addressed the question of whether our proposal leads to a radical rethinking of the relationship between the two.

In fact, there are two issues concerning [Figure R1](#). One is the extent to which production and comprehension are separate. In terms of the cognitive sandwich, are there two pieces of bread (rather than a wrap)? Our proposal is that instances of production involve comprehension processes ([Fig. 5](#)) and that instances of comprehension involve production processes ([Fig. 6](#)). But production and comprehension processes are nevertheless distinct – production involves mapping from intention to sound, and comprehension involves mapping from sound to intention. The second issue is whether the bread is separate from the filling. In other words, what is the relationship between production/comprehension and nonlinguistic mechanisms (thinking, general knowledge)? We propose that at least some aspects of general knowledge can be accessed during production and during comprehension, and moreover that interpreting the intention involves general knowledge. These aspects of general knowledge of course draw on a variety of cognitive functions such as memory and conflict resolution (see [Slevc & Novick](#)).

By interweaving production and comprehension, we have proposed that our account is incompatible with the traditional "cognitive sandwich" ([Hurley 2008a](#)) – an architecture in which production and comprehension are isolated from each other. Because production is a form of action, the use of production processes during comprehension means that comprehension involves a form of embodiment (i.e., uses action to aid perception). We also suggested that our account may be compatible with embodied accounts of meaning (e.g., [Barsalou 1999](#); [Glenberg & Gallese 2012](#)). In contrast, [Dove](#) argues that we assume different intermediate and disembodied levels of representation (e.g., syntax, phonology) that are not grounded in modality-specific input/output systems. We agree that our account is not compatible with this form of embodiment, and we accept his conclusion that we retain some amodal representations but abandon a form of modularity.

Our proposal for the relationship between production and comprehension runs counter to traditional interactive accounts which assume that cascading and feedback occur during production. For example, [Dell \(1986\)](#) assumed that a speaker activates semantic features and that activation cascades to words associated with those features and sounds associated with those words, and that activation then feeds back from the phonemes to the activated words and to other words involving those phonemes. Because this process involves several cycles before activation settles on a particular word and set of phonemes, [Dell](#) regards early stages of this process as involving prediction. In his very interesting proposal, cascading and feedback are internal to the implementer and are causal in bringing about the (implemented) linguistic representations. This also seems to be the position adopted by [Mylopoulos & Pereplyotchik](#), who replace our forward

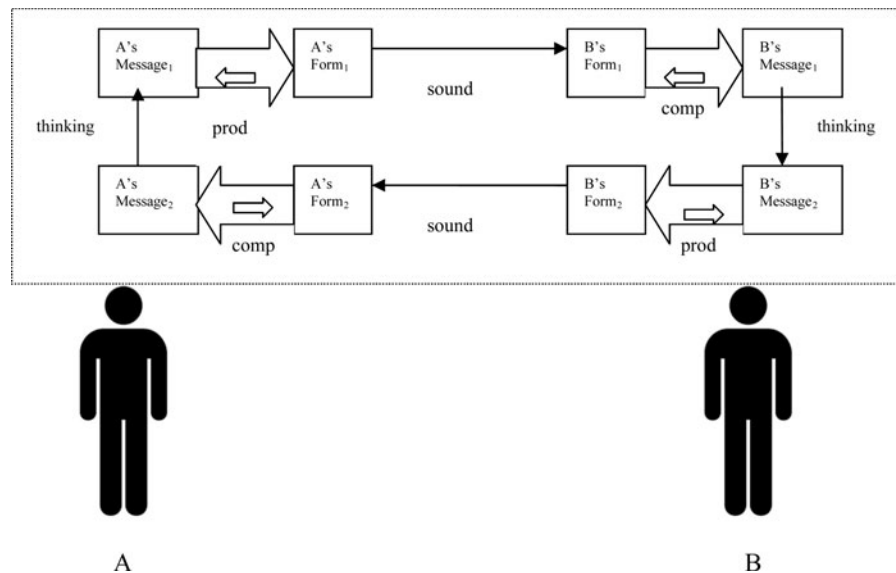


Figure R1. A traditional model of communication between A and B (repeated from target article).

model with an utterance plan internal to the implementer, and by **Mani & Huettig**, who regard it as a third route to prediction. In contrast, our predicted representations are the result of the efference copy of the production command, and they are therefore separate from the implementer. Our account of monitoring thus involves comparing two separate sets of representations and so is very different from both Dell and Mylopoulos & Pereplyotchik.

**Bowers'** discussion of Grossberg (1980) also appears similar to **Dell's** proposal and usefully demonstrates implementer-internal prediction. He also queries why forward modelling should speed up processing when its output is subsequently compared with the output of a slower implementer. The reason is that the comparison can be made as soon as the implementer's output is available (at any level). Otherwise, it is necessary to analyse the implementer's output (as in comprehension-based monitoring; Levelt 1989).

## R2. The neuroscience of production–comprehension relations

A number of commentators consider our account in relation to neuroscientific evidence. Some of this evidence concerns monitoring deficits associated with particular aphasia. **Hartsuiker** discusses a patient who cannot comprehend familiar sounds, words, or sentences, but who is nevertheless able to correct some of her own phonemic speech errors. He argues that such a patient would be incapable of monitoring her own speech by comparing the output of the implementer (utterance percept) with a forward model prediction (predicted utterance percept). However, it is difficult to draw clear conclusions from such cases without knowing the precise nature of the deficits. For example, this patient might monitor proprioceptively to correct errors. In relation to this, Tremblay et al. (2003) found that people can adapt their articulation to external perturbations of jaw movements during silently mouthed speech. In other words, they monitored and corrected their planned utterances in the absence of auditory

feedback. Hartsuiker also discusses a patient who detects phonemic errors in others but who does not repair his own (frequent) phonemic errors. It is possible that proprioceptive monitoring is disturbed (and that such monitoring is particularly important for repairing his phonemic errors), but outer-loop monitoring is preserved.

Other commentators point out that the classical Lichtheim–Broca–Wernicke neurolinguistic model is inconsistent with much recent neurophysiological data. In fact, we believe that their proposals are likely to be quite close to ours. **Hickok** assumes a dorsal stream which subserves sensorimotor integration for motor control (i.e., production) and a ventral stream which links sensory inputs to conceptual memory (i.e., comprehension). Both systems make use of prediction, with motor prediction facilitating production and sensory prediction facilitating comprehension. Hickok's position may not be so distinct from ours if we equate motor prediction with prediction-by-simulation and sensory prediction with prediction-by-association. However, unlike Hickok, we suggest that both streams may be called upon during comprehension. **Dick & Andric** also argue against the classical neurolinguistic model in favour of a dual-stream account. They suggest that the motor involvement in speech perception may only be apparent in perception under adverse conditions (see sect. R4 for a fuller discussion on this point). Finally, **Alario & Hamamé** point out additional evidence for forward modelling in production (e.g., Flinker et al. 2010). We accept that current evidence does not allow us to determine the neural basis of the efference copy, and we hope that this will be a target for future research.

## R3. Learning and development

Our target article did not explicitly discuss the role of forward modelling in language learning and development. However, we certainly recognise that our account is relevant to the acquisition of language production and comprehension, particularly in relation to their fluency. In fact, both forward and inverse models were first introduced

and tested as neurocomputational models of early skill acquisition (e.g., Jordan & Rumelhart 1992; Kawato et al. 1987; 1990), and we therefore argue that they can be applied to language. Accordingly, we are grateful to a number of commentators for fleshing out the importance of such models in the development of language. We also strongly agree that the use of forward and inverse models in learning does not disappear following childhood. Instead, it leads to adaptation and learning in adults, as well as to prediction of their own and others' utterances.

Hence, **Johnson, Turk-Browne, & Goldberg (Johnson et al.)** point out the role of prediction in learning to segment utterances, and in learning words and grammatical constructions. **Krishnan** notes various interrelations between production and comprehension abilities during development. We endorse her goal of explaining the mechanisms underlying such developmental changes. **Aitken** emphasizes the importance of the developmental perspective in accounting for communication processes in general, and we agree.

**Mani & Huettig** point out that two-year-olds' production vocabulary (rather than comprehension vocabulary) correlates with their ability to make predictions in a visual world situation (Mani & Huettig 2012). This provides important new evidence that prediction during comprehension makes use of production processes. In fact, it suggests that two-year-olds are already using prediction-by-simulation. Note that we speculated that adults may emphasize prediction-by-association when comprehending children; we did not suggest that children emphasize prediction-by-association during comprehension.

From a computational perspective, **Chang, Kidd, & Rowland (Chang et al.)** argue that linguistic prediction is a by-product of language learning. We accept that it originates in learning but note that it is critical for fluent performance in its own right. Their account of comprehension has some similarities to ours – in particular, that it uses a form of production-based prediction. However, it assumes separate meaning and sequencing pathways, whereas we adopt a more traditional multi-level account (semantics, syntax, phonology); future research could directly compare these accounts. We do not see why it is problematic to use syntax and semantics in supervised learning (any more than phonology, which is of course also abstract).

**McCauley & Christiansen** discuss a model of language acquisition in which prediction-by-simulation facilitates the model's shallow processing of the input during learning. They show how this model can account for a range of recent psycholinguistic findings about language acquisition. The integration of our account with the evidence for the representation of multi-word chunks is potentially informative. We also agree that shallow processing during comprehension may help explain apparent asymmetries between production and comprehension.

#### R4. Impoverished representations and production monitoring

Many commentators discuss our claim that predictions are impoverished. For example, **de Ruiter & Cummins** point out that Heinks-Maldonado et al.'s (2006) findings are compatible only with a forward model that incorporates

information about pitch. More generally, **Strijkers, Runqvist, Costa, & Holcomb (Strijkers et al.)** question how “poor” (predicted) representations can be used to successfully monitor “rich” (implemented) representations, and **Hartsuiker** similarly claims that impoverished representations are not a good standard for judging correctness (i.e., they will be particularly error-prone).

A key property of forward models is their flexibility. Their primary purpose (in adults) is to promote fluency, and therefore speakers are able to “tune into” whatever aspect of a stimulus is most relevant to this goal. So long as speakers know that pitch is relevant to a particular task (or is obviously being manipulated in an experiment), they predict the pitch that they will produce, and are disrupted if their predicted percept does not match the actual percept. People are able to determine what aspects of a percept to predict on the basis of their situation, such as the current experimental task (see **Howes, Healey, Eshghi, & Hough [Howes et al.]**) Such flexibility clearly makes the forward models more useful for aiding fluency, but it also means that we cannot determine which aspects of an utterance will necessarily be represented in a forward model. In **Alario & Hamamé's** terms, we assume that the “opt-out” is circumstantial rather than systematic. Hence, predictions may contain “fine-grained phonetic detail,” contra **Trude**.

In fact, the question of what information is represented in forward models of motor control (and learning) has received some attention. For example, Kawato et al. (1990) suggested that movement trajectories can be projected using critical *via-points* through which the trajectory has to pass at a certain time, rather than in terms of the moment-by-moment dynamics of the implemented movement trajectory. Optimal trajectories can then be learnt by applying local optimising principles for getting from one *via-point* to the next. In this way, impoverished predictions can be used to monitor rich implementations. In the same way, language users might predict particularly crucial aspects of an utterance, but the aspects that they predict will depend on the circumstances.

More generally, **Meyer & Hagoort** question the value of predicting one's own utterance. They argue that prediction is useful when it is likely to differ from the actual event. This is the case when predicting another person's behaviour or when the result of one's action is uncertain (e.g., moving in a strong wind). But they argue that people are confident about their own speech. Meyer & Hagoort admit that they will tend to be less confident in dialogue, and we agree. But more important, we argue that the behaviour of the production implementer is not fully determined by the production command, because the complex processes involved in production are subject to internal (“neural”) noise or priming (i.e., influences that may not be a result of the production command). Assuming that these sources of noise do not necessarily affect forward modelling as well, predicted speech may differ from actual speech. In addition, prediction is useful even if the behaviour is fully predictable, because it allows the actor to plan future behaviour on the basis of the prediction. In fact, we made such a proposal in relation to the order of heavy and light phrases (see sect. 3.1, target article).

**Hartsuiker** claims that our account incorrectly predicts an early competitor effect in Huettig and Hartsuiker (2010). His claim is based on the assumption that

comparing the predicted utterance percept with the actual utterance percept should involve phonological competitors in the same way that comprehending another's speech invokes such competitors. But the predicted utterance percept of *heart* does not also represent phonological competitors,<sup>1</sup> and the utterance percept is directly related to the predicted utterance percept (i.e., it is not analysed). Hartsuiker favours a conflict-monitoring account (e.g., Nozari et al. 2011), but such an account merely detects some difficulty during production, and it is unclear how it can determine the source of difficulty or the means of correction. We accept that a forward modelling account involves some duplication of information; a goal of our account and motor-control accounts is to provide reasons why complex biological systems are not necessarily parsimonious in this respect.

**Oppenheim** makes the interesting suggestion that inner speech might be the product of forward production models. This is an alternative to the possibility that inner speech involves an incomplete use of the implementer, in which the speaker inhibits production after computing a phonological or phonetic representation. Clearly, findings that inner speech is impoverished would provide some support for the forward-model account. But as Oppenheim himself notes, it is hard to see how inner slips could be identified without using the production implementer to generate an utterance percept at the appropriate level of representation (which is then compared to the predicted utterance percept).

**Jaeger & Ferreira** argue that the output of the forward production model (the predicted utterance) serves merely as input to the forward comprehension model, and suggest that the efference copy could directly generate the predicted utterance percept. In fact, the motivation for constructing the forward production model is to aid learning an inverse model that maps backward from the predicted utterance percept via the predicted utterance to the production command, just as in motor control theory (Wolpert et al. 2001). It is possible that sufficiently fluent speakers might be able to directly map from the production command to the predicted utterance percept (though there may be a separate mapping to the predicted utterance). But this would prevent speakers from remaining sufficiently flexible to learn new words or utterances, just as in the early stages of acquisition. In this context, Adank et al. (2010) found that adult comprehension of unfamiliar accents is facilitated by previous imitation of those accents to a similar extent whether speakers can or cannot hear their own speech. This suggests that adaptation is mediated by the forward production model (though there could also be an effect of proprioception). Note that the inverse model is not merely used for long-term learning, but is also used to modify an action as it takes place (e.g., to speak more clearly if background noise increases).

Finally, **Slevc & Novick** suggest that nonlinguistic memory tasks and linguistic conflict resolution involve common brain structures (the left inferior temporal gyrus). Patients with lesions to this area show difficulty with both memory tasks and with language production and comprehension. We propose that both self- and other-monitoring rely on memory-based predictions. One possibility is that monitoring can involve automatic correction when the difference between the prediction and the implementation is low, but monitoring requires extensive

access to general knowledge when there is greater discrepancy.

## R5. Prediction-by-simulation versus prediction-by-association

We propose that comprehenders make use of both prediction-by-simulation and prediction-by-association. In most situations, both routes provide some predictive value, and so we assume that comprehenders integrate the predictions that they make. Both routes also use domain-general cognitive mechanisms such as memory (see **Slevc & Novick**). We made some suggestions about when comprehenders are likely to weight one route more strongly than the other. For example, simulation will be weighted more strongly when the comprehender appears to be more similar to the speaker than otherwise. **Laurent, Moulin-Frier, Bessière, Schwartz, & Diard (Laurent et al.)** describe how they have modelled the contributions of association and simulation (in their terms, auditory and motor knowledge) to speech perception under both ideal and adverse conditions (both in the context of external noise and when the comprehender and speaker are very different). Under ideal conditions, association and simulation perform identically, but under adverse conditions, their performance falls off in different ways. However, they note that integration of the two routes (i.e., sensory-motor fusion) yields better performance, and we agree. We strongly support their programme of modelling these contributions (see also **de Ruiter & Cummins**) and agree that experiments conducted under adverse conditions may help discriminate the contributions of the two forms of prediction. We agree that their modelling results are consistent with findings from transcranial magnetic stimulation (TMS) studies which point to the contribution of motor systems to speech perception, but only in noisy conditions (e.g., D'Ausilio et al. 2011).

**Yoon & Brown-Schmidt** question the need for having both prediction-by-simulation and prediction-by-association in comprehension (i.e., dual-route prediction). These commentators claim that comprehenders using simulation would predict what the speaker would say (i.e., allocentrically). In fact, we propose that comprehenders use context to aid allocentric prediction, but that they are also subject to egocentric biases (i.e., comprehenders only partly take into account information about their partner's context). Yoon & Brown-Schmidt claim prediction-by-association would also be allocentric, and therefore question why we need two prediction mechanisms. We first note that prediction-by-association need not be allocentric, as it might be biased by prior perception of oneself.

But more important, the two routes to prediction are distinct for reasons unrelated to the role of context. Perhaps most important, prediction-by-simulation takes into account the inferred intention of the speaker in a way that prediction-by-association cannot (as it makes no reference to mental states). Hence, prediction-by-simulation should offer a richer and more situation-specific kind of prediction than prediction-by-association, and a combination of these predictions is likely to be more accurate than one form of prediction by itself.

To what extent can prediction-by-simulation account for speech adaptation effects? Trude and Brown-Schmidt

(2012) showed that listeners could use their knowledge of a speaker's regional pronunciation to rapidly rule out competitors in a visual world task (see also Dahan et al. 2008). **Trude** asks whether such rapid incorporation of context to aid prediction can be explained by prediction-by-simulation and to what extent it suggests that listeners make detailed phonetic predictions of what they will hear. As we proposed here (see sect. R4), forward models must be flexible, and in experimental situations that highlight detailed phonetic differences we would expect people to predict such details. Of course there is still the question of how rapidly a listener could incorporate the context (i.e., relating to speaker identity) into their forward model. However, it is interesting to note that Trude and Brown-Schmidt found that competitors were ruled out earlier following increased exposure to previous words (e.g., hearing *point to the bag* as opposed to just hearing *bag*). Trude also suggests that listeners' use of impoverished representations could explain difficulties in imitating those features. But, in fact, we claim that listeners typically compute fully specified representations using the implementer, so the reasons for difficulties in imitation presumably lie elsewhere. More generally, we claim that comprehenders use some production processes but not necessarily all (e.g., they may be unable to produce a particular accent).

**Festman** addresses issues relating to bilingualism. One reason for the difficulty of conversations between a native and a nonnative speaker is that their processing systems are likely to be very different (in terms of both speed and content) and so prediction-by-simulation is likely to be adversely affected. (In addition, prediction-by-simulation will be hindered by limited experience on which to learn forward models.) Prediction-by-association does not suffer from this problem. For example, most L1 (first-language) speakers have experience with L2 (second-language) speakers, and hence can predict L2 speakers even when they would behave differently from them. L2 speakers should also be able to predict L1 speakers (who they tend to encounter regularly), but they may of course not be able to make good predictions (e.g., if they do not know words that the L1 speakers would use).

**Rabagliati & Bemis** argue that much of language is not predictable and that its power is its ability to communicate the unpredictable. We do not claim that prediction underlies all of language comprehension. Rather, people use prediction whenever they can to assist comprehension (at different linguistic levels). But when an utterance is unpredictable, they simply rely on the implementer. In fact, failure to predict successfully serves to highlight the unexpected, and therefore allows the comprehender to concentrate resources.

## R6. Communicative intentions and the production command

Many of the commentators raise the issue of how our account is affected by communicative intentions. In our terms, this is the question of how the production command is determined and used. We agree with **Kashima, Bekkering, & Kashima (Kashima et al.)** that communicative intentions do not simply underlie the construction of semantics, syntax, and phonology, but incorporate information such as illocutionary force. Most

important, we do not assume that covert imitation simply involves copying linguistic representations (or that overt imitation involves "blind" repetition). Instead, our proposal (see Fig. 6, target article) is that comprehenders use the inverse model and context to derive the production command that comprehenders would use if they were to produce the speaker's utterance, and use this to drive the forward model (or to make overt responses). In other words, the forward model and overt imitation (or completion) are affected by the production command. Hence, our account is compatible with findings such as those of Ondobaka et al. (2011) because it proposes that imitation can be affected by aspects of intentions such as the compatibility between interlocutors' goals. It can also explain how accent convergence can depend on communicative intentions (e.g., relating to identity).

We agree with **Echterhoff** that our account aims to integrate a "language-as-action" approach to intention (represented as the production command) with the evidence for mechanistic time-locked processing. We limited our discussion of intentions for reasons of space, but agree that their relationship to other aspects of mental life is central to a more fully developed theory. He specifically highlights the importance of postdictive processing in determining nonliteral and other complex intentions, and we agree. We see his proposal as closely related to the use of offline prediction-by-simulation (see Pezzulo 2011a).

The HMOSAIC architecture is used to determine the relationship between actions and higher-level intentions. Commentators **de Ruiter & Cummins** query whether this is possible for language because of the particularly complex relationship between intention and utterance; **Pazzaglia** also claims that the mappings between intention and speech sounds are too complex for prediction-by-simulation. We are not convinced that this relationship is more complex than other aspects of human action (and interaction) and believe that this is simply an issue for future research.

**Kreysa** points out that comprehenders may use gaze to help predict utterances without deliberate consideration of intention. As she notes, such cues may constitute a form of prediction-by-association, though it is also possible that comprehenders perform prediction-by-simulation but with gaze constituting part of the context that is used to compute the intention (see Fig. 6, target article). If this is the case, gaze would help reduce the complexity of the intention-utterance relationship. She also questions whether anticipatory fixation in the visual-world paradigm involves prediction-by-association or simulation. In this context, we note that Lesage et al. (2012) showed that cerebellar repetitive TMS (rTMS) prevented such anticipatory fixations in predictive contexts (e.g., *The man will sail the boat*), but not in control sentences (or in vertex rTMS). As the cerebellum appears to be used for prediction in motor control (Miall & Wolpert 1996), this suggests that such fixations involve prediction-by-simulation.

**Jaeger & Ferreira** ask about the precise nature of the prediction errors that the system is trying to minimize. In particular, do these relate to evaluations of how well formed the output is or do they relate to evaluations of its communicative effectiveness in the context in which it is uttered? One possibility is that when speakers utter a predictable word, it means that they have forward modelled that word to a considerable extent before uttering it.

Hence, they weight the importance of the forward model more than the output of the implementer and therefore attenuate the form of the word. But alternatively, the speakers realize that the error tends to be less for predictable than unpredictable words, and know that addressees are less likely to comprehend words when the error is great. They thus use a strategy of clearer articulation when they realize that the error is likely to be great (perhaps based on their view of the addressee's ability to comprehend).

**Pezzulo & Dindo** argue that producers use intentional (signalling) strategies to aid their comprehenders' predictions. In other words, they make themselves more predictable to the comprehender. To do this, they maintain an internal model of the comprehenders' uncertainty. Hence, the authors propose a type of allocentric account. We suspect that speakers make their utterances predictable for both allocentric reasons (e.g., lengthening vowels for children) and egocentric reasons. For example, an effect of interactive alignment (Pickering & Garrod 2004) is to make interlocutors more similar (see sect. 3.3, target article) so that comprehenders are more likely to predict speakers accurately. Such alignment might itself follow from an intentional strategy to align but need not. In their penultimate paragraph, Pezzulo & Dindo make several very interesting additional suggestions about ways in which producers and comprehenders may make use of prediction beyond those we addressed in the target article.

## R7. Interleaving production and comprehension in dialogue

**Howes et al.** criticise traditional models of production and comprehension based on large units (such as whole sentences), and we agree. Such models are clearly unable to deal with the fragmentary nature of many contributions to dialogue. Howes et al. propose an incremental model that combines production and comprehension. This model may be able to deal with incrementality and joint utterances in dialogue but does not provide an account of prediction, either of upcoming words in monologue (e.g., DeLong et al. 2005) or ends of turns in dialogue (de Ruiter et al. 2006). Moreover, Howes et al. argue that prediction can only help speakers repeat their interlocutors, but this is not the case. Overt responses based on the derived production command for the current utterance ( $i_B(t)$ ) lead to repetition, but overt responses based on the derived production command for the upcoming utterance ( $i_B(t+1)$ ) lead to continuations (see Fig. 6, target article). In other words, we believe that our account provides mechanisms that underlie dialogue (Pickering & Garrod 2004).

**Fowler** argues that we wrongly emphasize internal predictive models when the same benefits can be accrued by directly interacting with the environment, most importantly our interlocutors. We accept that the information in the environment helps determine people's actions, but argue that predictions driven by internal models that have been shaped by past experience allow people to perform better (just as is the case for complex engineering). In the interaction process, overt imitation, completions, and complementary responses all appear to occur regularly, and all are compatible with our account (see sect. 3.3, target article).

**Echterhoff** first asks whether our action-based account can generalize to noninteractive situations. Our models of production and comprehension are explained noninteractively and then applied to dialogue. In particular, we propose that dialogue allows interlocutors to make use of overt responses; in monologue, such overt responses are not relevant, therefore language users focus more completely on internal processes. Echterhoff also argues that shared reality (see sect. 2.3, target article) involves more than seamless coordinated activity, but also has an evaluative component. We propose that successful mutual prediction (both A and B correctly predict both A and B) underlies shared reality and is in turn likely to support the alignment of evaluations of that activity. Mutual predictions occur as a result of aligned action commands, and action commands reflect intentions which of course involve the motivations that, according to Echterhoff, underlie shared reality.

## ACKNOWLEDGMENTS

We thank Martin Corley and Chiara Gambi for their comments and acknowledge support of ESRC Grants no. RES-062-23-0736 and no. RES-060-25-0010.

## NOTE

1. This constitutes a sense in which the predicted utterance percept is impoverished with respect to the utterance percept, but does not lose any relevant information about the utterance.

## References

[The letters "a" and "r" before author's initials stand for target article and response references, respectively]

- Acheson, D. J. & MacDonald, M. C. (2009) Verbal working memory and language production: Common approaches to the serial ordering of verbal information. *Psychological Bulletin* 135(1):50–68. [LRS]
- Adank, P. & Devlin, J. T. (2010) On-line plasticity in spoken sentence comprehension: Adapting to time-compressed speech. *NeuroImage* 49:1124–32. [aMJJP]
- Adank, P., Hagoort, P. & Bekkering, H. (2010) Imitation improves language comprehension. *Psychological Science* 21:1903–909. [arMJP]
- Aglioti, S. M. & Pazzaglia, M. (2010) Representing actions through their sound. *Experimental Brain Research* 206(2):141–51. DOI: 10.1007/s00221-010-2344-x. [MP]
- Aglioti, S. M. & Pazzaglia, M. (2011) Sounds and scents in (social) action. *Trends in Cognitive Sciences* 15(2):47–55. DOI: 10.1016/j.tics.2010.12.003. [MP]
- Aitken, K. J. (2008) Intersubjectivity, affective neuroscience, and the neurobiology of autistic spectrum disorders: A systematic review. *Keio Journal of Medicine* 57:15–36. [KJA]
- Aitken, K. J. & Trevarthen, C. (1997) Self/other organization in human psychological development. *Development and Psychopathology* 9:653–77. [KJA]
- Alcock, K. J. & Krawczyk, K. (2010) Individual differences in language development: Relationship with motor skill at 21 months. *Developmental Science* 13(5):677–91. [SK]
- Allopena, P. D., Magnuson, J. S. & Tanenhaus, M. K. (1998) Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language* 38:419–39. [aMJJP]
- Altmann, G. T. M. (2011) The mediation of eye movements by spoken language. In *The Oxford handbook of eye movements*, ed. S. P. Liversedge, I. D. Gilchrist & S. Everling, pp. 979–1003. Oxford University Press. [HK]
- Altmann, G. T. M. & Kamide, Y. (1999) Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition* 73(3):247–64. [FC, HK, aMJJP, LRS]
- Altmann, G. T. M. & Mirkovic, J. (2009) Incrementality and prediction in human sentence processing. *Cognitive Science* 33(4):583–609. [aMJJP, HR]
- Anders, S., Heinze, J., Weiskopf, N., Ethofer, T. & Haynes, J.-D. (2011) Flow of affective information between communicating brains. *NeuroImage* 54:439–46. [KJA]

- Anderson, M. L. (2003) Embodied cognition: A field guide. *Artificial Intelligence* 149, 151–56. [GD]
- Andric, M. & Small, S. L. (2012) Gesture's neural language. *Frontiers in Psychology* 3:99. DOI: 10.3389/fpsyg.2012.00099. [ASD]
- Arbib, M. A. (2005) From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences* 28(2):105–24. [MP]
- Arbib, M. A. (2012) *How the brain got language: The mirror system hypothesis*. Oxford University Press. [KJA]
- Arnon, I. & Snider, N. (2010) More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language* 62:67–82. [SMM]
- Aronoff, M. (1976) Word formation in generative grammar. *Linguistic Inquiry Monograph 1*. MIT Press. [MAJ]
- Austin, J. L. (1962) *How to do things with words*. Clarendon Press. [GE]
- Aylett, M., & Turk, A. (2004) The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech* 47(1):31–56. [TFJ]
- Baayen, R. H., Milin, P., Filipović Durdević, D., Hendrix, P. & Marelli, M. (2011) An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review* 118:438–82. [TFJ]
- Babel, M. (2010) Dialect divergence and convergence in New Zealand English. *Language in Society* 39:437–56. [YK]
- Babel, M. (2012) Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics* 40:177–89. [AMT]
- Bannard, C. & Matthews, D. (2008) Stored word sequences in language learning: The effect of familiarity on children's repetition of four-word combinations. *Psychological Science* 19:241–48. [SMM]
- Bar, M. (2009) The proactive brain: Memory for predictions. *Philosophical Transactions of the Royal Society B* 364:1235–43. [LRS]
- Bargh, J. A. & Chartrand, T. L. (1999) The unbearable automaticity of being. *American Psychologist* 54:462–79. [GP]
- Barrett, J. & Fleming, A. S. (2011) Annual research review: All mothers are not created equal: Neural and psychobiological perspectives on mothering and the importance of individual differences. *Journal of Child Psychology and Psychiatry* 52:368–97. [KJA]
- Barsalou, L. (1999) Perceptual symbol systems. *Behavioral and Brain Sciences* 22:577–600. [arMJP]
- Barsalou, L. W. (2008) Grounded cognition. *Annual Review of Psychology* 59:617–45. DOI 10.1146/annurev.psych.59.103006.093639. [GD, MP]
- Barsalou, L. W., Simmons, W. K., Barbey, A. K. & Wilson, C. D. (2003) Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Science* 7:84–91. [GD]
- Bates, E., Benigni, L., Bretherton, I., Camaioni, L. & Volterra, V. (1979) *The emergence of symbols: Cognition and communication in infancy*. Academic Press. [SK]
- Bates, E., Bretherton, I. & Snyder, L. (1988) *From first words to grammar: Individual differences and dissociable mechanisms*. Cambridge University Press. [SK]
- Bates, E. & Dick, F. (2002) Language, gesture, and the developing brain. *Developmental Psychobiology* 40(3):293–310. [SK, MP]
- Bavelas, J. B., Coates, L. & Johnson, T. (2000) Listeners as co-narrators. *Journal of Personality and Social Psychology* 79:941–52. [aMJP]
- Bedny, M. & Caramazza, A. (2011) Perception, action, and word meanings in the human brain: The case from action verbs. *Annals of the New York Academy of Sciences* 1224:81–95. DOI:10.1111/j.1749-6632.2011.06013.x. [ASD]
- Beier, J. S. & Spelke, E. S. (2012) Infants' developing understanding of social gaze. *Child Development* 83:486–96. [KJA]
- Ben Shalom, D. & Poeppel, D. (2008) Functional anatomic models of language: assembling the pieces. *The Neuroscientist* 14:119–27. [KJA, aMJP]
- Bencini, G. M. & Goldberg, A. E. (2000) The contribution of argument structure constructions to sentence meaning. *Journal of Memory & Language* 43:640–51. [MAJ]
- Bessière, P., Laugier, C. & Siegwart, R. ed. (2008) *Probabilistic reasoning and decision making in sensory-motor systems, volume 46 of Springer tracts in advanced robotics*. Springer-Verlag. [RL]
- Binder, J. R., Desai, R. H., Graves, W. W. & Conant, L. L. (2009) Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex* 19:2767–96. [ASD]
- Blackmer, E. R. & Mitton, J. L. (1991) Theories of monitoring and the timing of repairs in spontaneous speech. *Cognition* 39:173–94. [aMJP]
- Blakemore, S.-J., Frith, C. D. & Wolpert, D. M. (1999) Spatio-temporal prediction modulates the perception of self-produced stimuli. *Journal of Cognitive Neuroscience* 11:551–59. [aMJP]
- Boatman, D., Gordon, B., Hart, J., Selnes, O., Miglioretti, D. & Lenz, F. (2000) Transcortical sensory aphasia: revisited and revised. *Brain* 123(8):1634–42. [HR]
- Bock, J. K. (1996) Language production: Methods and methodologies. *Psychonomic Bulletin & Review* 3:395–421. [aMJP]
- Bock, J. K. & Levelt, W. J. M. (1994) Language production: Grammatical encoding. In: *Handbook of psycholinguistics*, ed. M. A. Gernsbacher, Academic Press. [aMJP]
- Bock, K. & Miller, C. A. (1991) Broken agreement. *Cognitive Psychology* 23:45–93. [aMJP]
- Borensztajn, G., Zuidema, J. & Bod, R. (2009) Children's grammars grow more abstract with age. *Topics in Cognitive Science* 1:175–88. [SMM]
- Borovsky, A., Elman, J. L. & Fernald, A. (2012) Knowing a lot for one's age: Vocabulary skill and not age is associated with anticipatory incremental sentence interpretation in children and adults. *Journal of Experimental Child Psychology* 112(4):417–36. [FC]
- Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S. & Cohen, J. D. (2001) Conflict monitoring and cognitive control. *Psychological Review* 108:624–52. [RJH]
- Bourhis, R. Y. & Giles, H. (1977) The language of intergroup distinctiveness. In: *Language, ethnicity, and intergroup relations*, ed. H. Giles, pp. 119–36. Academic Press. [YK]
- Boyd, J. K. & Goldberg, A. E. (2011) Learning what not to say: Categorization and statistical preemption in “a-adjective” production. *Language* 87:1–29. [MAJ]
- Branigan, H. P., Pickering, M. J. & Cleland, A. A. (2000) Syntactic coordination in dialogue. *Cognition* 75:B13–25. [aMJP]
- Branigan, H. P., Pickering, M. J. & McLean, J. F. (2005) Priming prepositional-phrase attachment during language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 31:468–81. [LRS]
- Bråten, S. (1998) *Intersubjective communication and emotion in early ontogeny*. Cambridge University Press. [KJA]
- Bratman, M. E. (1992) Shared cooperative activity. *The Philosophical Review* 101:327–41. [YK]
- Bratman, M. E. (1999) *Faces of intention: Selected essays on intention and agency*. Cambridge University Press. [YK]
- Brennan, S. E. & Clark, H. H. (1996) Conceptual pacts and lexical choice in conversation. *Learning, Memory* 22(6):1482–93. [TFJ]
- Brooks, P. J. & Tomasello, M. (1999) How children constrain their argument structure constructions. *Language* 75(4):720–38. [MAJ]
- Brown, R. & McNeill, D. (1966) The “tip of the tongue” phenomenon. *Journal of Verbal Learning and Verbal Behavior* 5:325–37. [aMJP]
- Brown-Schmidt, S. (2009) The role of executive function in perspective taking during online language comprehension. *Psychonomic Bulletin & Review* 16(5):893–900. [LRS]
- Brown-Schmidt, S. & Tanenhaus, M. K. (2008) Real-time investigation of referential domains in unscripted conversation: A targeted language game approach. *Cognitive Science* 32:643–84. [SOY]
- Buccino, G., Riggio, L., Melli, G., Binkofski, F., Gallese, V. & Rizzolatti, G. (2005) Listening to action-related sentences modulates the activity of the motor system: A combined TMS and behavioral study. *Brain Research and Cognition Brain Research* 24(3):355–63. DOI: 10.1016/j.cogbrainres.2005.02.020. [MP]
- Buxbaum, L. J., Kyle, K. M. & Menon, R. (2005) On beyond mirror neurons: Internal representations subserving imitation and recognition of skilled object-related actions in humans. *Brain Research and Cognition Brain Research* 25(1):226–39. DOI: 10.1016/j.cogbrainres.2005.05.014. [MP]
- Callan, D. E., Jones, J. A., Callan, A. M. & Akahane-Yamada, R. (2004) Phonetic perceptual identification by native- and second-language speakers differentially activates brain regions involved with acoustic phonetic processing and those involved with articulatory-auditory/orosensory internal models. *NeuroImage* 22:1182–94. [ASD]
- Cappa, S. F. & Pulvermüller, F. (2012) Cortex special issue: Language and the motor system. *Cortex* 48(7):785. [ASD]
- Carpenter, W. B. (1852) On the influence of suggestion in modifying and directing muscular movement independently of volition. *Proceedings of the Royal Institution* 147–54. [YK]
- Castelli, L., Pavan, G., Ferrari, E. & Kashima, Y. (2009) The stereotyper and the chameleon: The effects of stereotype use on perceiver's mimicry. *Journal of Experimental Social Psychology* 4:835–39. [YK]
- Castellini, C., Badino, L., Metta, G., Sandini, G., Tavella, M., Grimaldi, M. & Fadiga, L. (2011) The use of phonetic motor invariants can improve automatic phoneme discrimination. *PLoS ONE* 6(9):e24055. [RL]
- Catani, M. & ffytche, D. H. (2005) The rises and falls of disconnection syndromes. *Brain* 128:2224–39. [KJA]
- Catmur, C., Walsh, V. & Heyes, C. (2007) Sensorimotor learning configures the human mirror system. *Current Biology* 17(17):1527–31. [GH]
- Chandrasekaran, C., Trubanova, A., Stillitano, S., Caplier, A. & Ghazanfar, A. A. (2009) The natural statistics of audiovisual speech. *PLoS Computational Biology* 5(7):e1000436. DOI:10.1371/journal.pcbi.1000436. [ASD]

- Chang, F. (2002) Symbolically speaking: A connectionist model of sentence production. *Cognitive Science* 26(5):609–51. [FC]
- Chang, F. (2009) Learning to order words: A connectionist model of heavy NP shift and accessibility effects in Japanese and English. *Journal of Memory and Language* 61(3):374–97. [FC]
- Chang, F., Dell, G. S. & Bock, K. (2006) Becoming syntactic. *Psychological Review* 113(2):234–272. [FC, TF, NM, SMM, aMJP, LRS]
- Chartrand, T. L. & Bargh, J. A. (1999) The chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology* 76:893–910. [aMJP]
- Chemero, A. (2009) *Radical embodied cognitive science*. MIT Press. [GD]
- Clark, A. (2013) Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Brain and Behavioral Sciences* 36(3):181–253. [TFJ]
- Clark, H. H. (1996) *Using language*. Cambridge University Press. [YK, GP, aMJP, LRS, SOY]
- Clark, H. H. & Wilkes-Gibbs, D. (1986) Referring as a collaborative process. *Cognition* 22:1–39. [GE, aMJP]
- Clayards, M., Tanenhaus, M. K., Aslin, R. N. & Jacobs, R. A. (2008) Perception of speech reflects optimal use of probabilistic speech cues. *Cognition* 108(3):804–809. [TFJ]
- Cohn, J. F. & Tronick E. Z. (1988) Mother-infant face-to-face interaction: Influence is bidirectional and unrelated to periodic cycles in either partner's behavior. *Developmental Psychology* 24:386–92. [KJA]
- Colas, F., Diard, J. & Bessi re, P. (2010) Common Bayesian models for common cognitive issues. *Acta Biotheoretica* 58(2–3):191–216. [RL]
- Condon, W. S. & Sander, L. W. (1974) Neonate movement is synchronized with adult speech: Interactional participation and language acquisition. *Science* 183:99–101. [KJA]
- Conway, C. M., Bauernschmidt, A., Huang, S. S. & Pisoni, D. B. (2010) Implicit statistical learning in language processing: Word predictability is the key. *Cognition* 114:356–71. [MAJ]
- Corballis, M. C. (2009) The evolution of language. *Annals of the New York Academy of Sciences* 1156:19–43. DOI: 10.1111/j.1749-6632.2009.04423.x. [MP]
- Corley, M., Brocklehurst, P. H. & Moat, H. S. (2011) Error biases in inner and overt speech: Evidence from tongue twisters. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 37(1):162–75. DOI:10.1037/a0021321. [GMO, aMJP]
- Cowles, H. W., Walenski, M. & Kluender, R. (2007) Linguistic and cognitive prominence in anaphor resolution: Topic, contrastive focus and pronouns. *Topoi* 26:3–18. [LRS]
- Crain, S. & Fodor, J. D. (1985) How can grammars help parsers? In: *Natural language parsing: Psychological, computational, and theoretical perspectives*, ed. D. R. Dowty, L. Karttunen & A. M. Zwicky, pp. 94–128. Cambridge University Press. [GH]
- Csibra, G. & Gergely, G. (2007) 'Obsessed with goals': Functions and mechanisms of teleological interpretation of actions in humans. *Acta Psychologica* 124:60–78. [aMJP]
- Cubelli, R., Marchetti, C., Boscolo, G. & Della Sala, S. (2000) Cognition in action: Testing a model of limb apraxia. *Brain and Cognition* 44(2):144–65. DOI: 10.1006/brcg.2000.1226. [MP]
- Cutting, J. C. & Ferreira, V. S. (1999) Semantic and phonological information flow in the production lexicon. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 25:318–44. [GSD]
- D'Ausilio, A., Badino, L., Li, Y., Tokay, S., Craighero, L., Canto, R., Aloimonos, Y. & Fadiga, L. (2012a) Leadership in orchestra emerges from the causal relationships of movement kinematics. *PLoS ONE* 7(5): e35757. DOI:10.1371/journal.pone.0035757. [GP]
- D'Ausilio, A., Bufalari, I., Salmas, P. & Fadiga, L. (2012b) The role of the motor system in discriminating degraded speech sounds. *Cortex* 48:882–87. [RL]
- D'Ausilio, A., Jarmolowska, J., Busan, P., Bufalari, I. & Craighero, L. (2011) Tongue corticospinal modulation during attended verbal stimuli: Priming and coarticulation effects. *Neuropsychologia* 49:3670–76. [arMJP]
- D'Ausilio, A., Pulverm ller, F., Salmas, P., Bufalari, I., Begliomini, C. & Fadiga, L. (2009) The motor somatopy of speech perception. *Current Biology* 19:381–85. [KJA, ASD, aMJP]
- Dahan, D., Drucker, S. J. & Scarborough, R. A. (2008) Talker adaptation in speech perception: Adjusting the signal or the representations? *Cognition* 108:710–18. [rMJP, AMT]
- Damasio, A. R. & Damasio, H. (1994) Cortical systems for retrieval of concrete knowledge: The convergence zone framework. In: *Large-scale neuronal theories of the brain. Computational neuroscience*, ed. C. Koch & J. L. Davis, pp. 61–74. MIT Press. [GD]
- Davidson, P. R. & Wolpert, D. M. (2005) Widespread access to predictive models in the motor system: A short review. *Journal of Neural Engineering* 2:S13–19. [aMJP]
- De Ruiter, J. P. & Cummins, C. (2012) *A model of intentional communication: AIRBUS (Asymmetric Intention Recognition with Bayesian Updating of Signals)*. In: *Proceedings of SemDial 2012*, ed. S. Brown-Schmidt, J. Ginzburg & S. Larsson, pp. 149–50. [JPdR]
- De Ruiter, J. P., Mitterer, H. & Enfield, N. J. (2006) Predicting the end of a speaker's turn: a cognitive cornerstone of conversation. *Language* 82(3):515–35. [JPdR, arMJP]
- Delaherche, E., Chetouani, M., Mahdhaoui, A., Saint-Georges, C., Viaux, S. & Cohen, D. (2012) Interpersonal synchrony: A survey of evaluation methods across disciplines. *IEEE Transactions on Affective Computing* 3(3): 349–65. [KJA]
- Dell, G. S. (1978) Slips of the mind. In: *The fourth Lacus forum*, ed. M. Paradis, pp. 69–75. Hornbeam Press. [GMO]
- Dell, G. S. (1986) A spreading-activation theory of retrieval in sentence production. *Psychological Review* 93:283–321. [GSD, JPdR, arMJP]
- Dell, G. S. (1988) The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of Memory and Language* 27:124–42. [aMJP]
- Dell, G. S., Burger, L. K. & Svec, W. R. (1997) Language production and serial order: A functional analysis and a model. *Psychological Review* 104(1):123–47. [LRS]
- Dell, G. S. & Repka, R. J. (1992) Errors in inner speech. In: *Experimental slips and human error: Exploring the architecture of volition*, ed. B. J. Baars, pp. 237–62. Plenum. [GMO]
- Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M. & Gagnon, D. A. (1997) Lexical access in aphasic and nonaphasic speakers. *Psychological Review* 104:801–38. [GSD, aMJP]
- DeLong, K. A., Urbach, T. P. & Kutas, M. (2005) Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience* 8(8):1117–21. [NM, arMJP, HR]
- Denton, D. (2005) *The primordial emotions: The dawning of consciousness*. Oxford University Press. [KJA]
- Desai, R. H., Binder, J. R., Conant, L. L. & Seidenberg, M. S. (2010) Activation of sensory-motor areas in sentence comprehension. *Cerebral Cortex* 20:468–78. [aMJP]
- DeVault, D., Sagae, K. & Traum, D. (2011) Incremental interpretation and prediction of utterance meaning for interactive dialogue. *Dialogue and Discourse* 2(1):143–70. [JPdR]
- Di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V. & Rizzolatti, G. (1992) Understanding motor events: A neurophysiological study. *Experimental Brain Research* 91(1):176–80. [MP, aMJP]
- Dick, A. S., Goldin-Meadow, S., Hasson, U., Skipper, J. I. & Small, S. L. (2009) Co-speech gestures influence neural activity in brain regions associated with processing semantic information. *Human Brain Mapping* 30:3509–26. [ASD]
- Dick, A. S., Goldin-Meadow, S., Solodkin, A. & Small, S. L. (2012) Gesture in the developing brain. *Developmental Science* 15:165–80. [ASD]
- Dick, A. S., Solodkin, A. & Small, S. L. (2010) Neural development of networks for audiovisual speech comprehension. *Brain and Language* 114:101–14. [ASD]
- Dick, A. S. & Tremblay, P. (2012) Beyond the arcuate fasciculus: Consensus and controversy in the connectome of language. *Brain: A Journal of Neurology* 135:3529–50. doi: 10.1093/brain/aww222. [ASD]
- Dijksterhuis, A. & Bargh, J. A. (2001) The perception-behavior expressway: Automatic effects of social perception on social behavior. In: *Advances in experimental social psychology*, vol. 33, ed. M. P. Zanna, pp. 1–40. Academic Press. [aMJP]
- Dikker, S. & Pykk nen, L. (2011) Before the N400: Effects of lexical-semantic violations in visual cortex. *Brain and Language* 118:23–28. [aMJP]
- Dikker, S., Rabagliati, H., Farmer, T. A. & Pykk nen, L. (2010) Early occipital sensitivity to syntactic category is based on form typicality. *Psychological Science* 21:629–34. [aMJP]
- Dikker, S., Rabagliati, H. & Pykk nen, L. (2009) Sensitivity to syntax in visual cortex. *Cognition* 110(3):293–321. [aMJP, HR]
- Dindo, H., Zambuto, D. & Pezzulo, G. (2011) Motor simulation via coupled internal models using sequential Monte Carlo. *Proceedings of IJCAI 2011*:2113–19. [GP]
- Dodd, B. & McIntosh, B. (2010) Two-year-old phonology: Impact of input, motor and cognitive abilities on development. *Journal of Child Language* 37(5):1027–46. [SK]
- Doucet, S., Soussignan, R., Sagot, P. & Schaal, B. (2009) The secretion of areolar (montgomery's) glands from lactating women elicits selective, unconditional responses in neonates. *PLoS ONE*, 4(10):e7579. DOI:10.1371/journal.pone.0007579. [KJA]
- Echterhoff, G., Higgins, E. T., Kopietz, R. & Groll, S. (2008) How communication goals determine when audience tuning biases memory. *Journal of Experimental Psychology: General* 137:3–21. [GE]
- Echterhoff, G., Higgins, E. T. & Levine, J. M. (2009) Shared reality: Experiencing commonality with others' inner states about the world. *Perspectives on Psychological Science* 4:496–521. [GE, aMJP]
- Eickhoff, S. B., Heim, S., Zilles, K. & Amunts, K. (2009) A systems perspective on the effective connectivity of overt speech production. *Philosophical*

- Transactions of the Royal Society A: Mathematical, Physical, and Engineering Sciences* 367(1896):2399–421. DOI:10.1098/rsta.2008.0287. [ASD]
- Elman, J. L. (1990) Finding structure in time. *Cognitive Science*, 14(2), 179–211. [FC, aMJP]
- Elman, J. L. (1991) Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning* 7:195–225. [MAJ]
- Elman, J. L. (1993) Learning and development in neural networks: The importance of starting small. *Cognition* 48:71–99. [MAJ]
- Emery, N. J. (2000) The eyes have it: The neuroethology, function and evolution of social gaze. *Neuroscience and Biobehavioral Reviews* 24:581–604. [HK]
- Erten, I. H. & Razi, S. (2009) The effects of cultural familiarity on reading comprehension. *Reading in a Foreign Language* 21:60–77. [JF]
- Evans, N. & Levinson, S. C. (2009) The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences* 32(5):429–92. [NM]
- Fadiga, L. & Craighero, L. (2006) Hand actions and speech representation in Broca's area. *Cortex* 42(4):486–490. [MP]
- Fadiga, L., Craighero, L., Buccino, G. & Rizzolatti, G. (2002) Speech listening specifically modulates the excitability of tongue muscles: A TMS study. *European Journal of Neuroscience* 15:399–402. [aMJP]
- Falk, D. (2004) Prelinguistic evolution in early hominins: Whence motherese? *Behavioral and Brain Sciences* 27:491–503, commentaries 503–41. [KJA]
- Farmer, T. A., Brown, M. & Tanenhaus, M. K. (2013) Prediction, explanation, and the role of generative models in language processing. *Brain and Behavioral Sciences* 36(3):211–12. [TFJ]
- Federmeier, K. D. (2007) Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology* 44:491–505. [aMJP, KS]
- Federmeier, K. D. & Kutas, M. (1999) A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language* 41:469–95. [SOY]
- Federmeier, K., McLennan, D. B. & De Ochoa, E. & Kutas, M. (2002) The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An ERP study. *Psychophysiology*, 39:133–46. [NM]
- Fedorenko, E., Gibson, E. & Rohde, D. (2006) The nature of working memory capacity in sentence comprehension: Evidence against domain-specific working memory resources. *Journal of Memory & Language* 54:541–53. [LRS]
- Fedzechkina, M., Jaeger, T. F. & Newport, E. (2012) Language learners restructure their input to facilitate efficient communication. *Proceedings of the National Academy of Sciences of the United States of America* 109(44):17897–902. [TFJ]
- Feldman, R. (2007) On the origins of background emotions: From affect synchrony to symbolic expression. *Emotion* 7:601–11. [KJA]
- Ferguson, H. J., Scheepers, C. & Sanford, A. J. (2010) Expectations in counterfactual and theory of mind reasoning. *Language and Cognitive Processes* 25:297–346. [SOY]
- Ferreira, F. (2003) The misinterpretation of noncanonical sentences. *Cognitive Psychology* 47:164–203. [aMJP]
- Ferreira, F., Ferraro, V. & Bailey, K. G. D. (2002) Good enough representations in language comprehension. *Current Directions in Psychological Science* 11:11–15. [SMM]
- Ferreira, F. & Tanenhaus, M. (2007) Introduction to the special issue on language-vision interactions. *Journal of Memory and Language* 57:455–59. [CAF]
- Ferreira, V. S. (1996) Is it better to give than to donate? Syntactic flexibility in language production. *Journal of Memory and Language* 35:724–55. [aMJP]
- Ferreira, V. S. (2008) Ambiguity, accessibility, and a division of labor for communicative success. *Psychology of Learning and Motivation* 49:209–46. [TFJ]
- Fine, A. B. & Jaeger, T. F. (2013) Evidence for implicit learning in syntactic comprehension. *Cognitive Science* 37(3):578–91. [TFJ]
- Fine, A. B., Jaeger, T. F., Farmer, T. & Qian, T. (submitted) Rapid linguistic adaptation during syntactic comprehension. [TFJ]
- Fischer, M. H. & Zwaan, R. A. (2008) Embodied language: A review of the role of the motor system in language comprehension. *Quarterly Journal of Experimental Psychology* (2006) 61(6):825–50. DOI:10.1080/17470210701623605. [ASD, aMJP]
- Fitch, W. T. (2006) The biology and evolution of music: A comparative perspective. *Cognition* 100:173–215. [KJA]
- Fivaz-Depeursinge, E. & Favez, N. (2006) Exploring triangulation in infancy: Two contrasted cases. *Family Process* 45:3–18. [KJA]
- Flinker A., Chang E. F., Kirsch H. E., Barbaro N. M., Crone, N. E. & Knight R. T. (2010) Single-trial speech suppression of auditory cortex activity in humans. *Journal of Neuroscience* 30:16643–50. [F-XA, rMJP]
- Fodor, J. A. (1983) *The modularity of mind*. MIT Press. [aMJP]
- Fogassi, L., Ferrari, P. F., Gesierich, B., Rozzi, S., Chersi, F. & Rizzolatti, G. (2005) Parietal lobe: From action organization to intention understanding. *Science* 308(5722):662–67. DOI: 10.1126/science.1106138. [MP]
- Fontana, A. P., Kilner, J. M., Rodrigues, E. C., Joffily, M., Nighoghossian, N., Vargas, C. D. & Sirigu, A. (2012) Role of the parietal cortex in predicting incoming actions. *NeuroImage* 59(1):556–64. DOI: 10.1016/j.neuroimage.2011.07.046. [MP]
- Fowler, C. A., Brown, J., Sabadini, L. & Weihing, J. (2003) Rapid access to speech gestures in perception: Evidence from choice and simple response time tasks. *Journal of Memory and Language* 49:296–314. [aMJP]
- Frank, A. (2011) *Integrating linguistic, motor, and perceptual information in language production*. University of Rochester. [TFJ]
- Fraser, C., Bellugi, U. & Brown, R. (1963) Control of grammar in imitation, comprehension, and production. *Journal of Verbal Learning and Verbal Behavior* 2:121–35. [SMM]
- Frazier, L. (1987) Sentence processing: A tutorial review. In: *Attention and performance XII: The psychology of reading*, ed. M. Coltheart, pp. 559–86. Erlbaum. [aMJP]
- Frazier, L. & Flores d'Arcais, G. (1989) Filler driven parsing: A study of gap filling in Dutch. *Journal of Memory and Language* 28:331–44. [GH]
- French, R. M., Addyman, C. & Mareschal, D. (2011) TRACX: A recognition-based connectionist framework for sequence segmentation and chunk extraction. *Psychological Review* 118:614–636. [MAJ]
- Freudenthal, D., Pine, J. M., Aguado-Orea, J. & Gobet, F. (2007) Modelling the developmental pattern of finiteness marking in English, Dutch, German, and Spanish using MOSAIC. *Cognitive Science* 31:311–41. [SMM]
- Frey S., Campbell J. S., Pike G. B. & Petrides M. (2008) Dissociating the human language pathways with high angular resolution diffusion fiber tractography. *Journal of Neuroscience* 28:11435–44. [F-XA]
- Freyd, J. J. & Finke, R. A. (1984) Representational momentum. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 10:126–32. [aMJP]
- Friston, K. J., Daunizeau, J., Kilner, J. & Kiebel, S. J. (2010) Action and behavior: A free-energy formulation. *Biological Cybernetics* 102(3):227–60. [GH]
- Frith, C. D. & Frith, U. (2008) Implicit and explicit processes in social cognition. *Neuron* 60(3):503–10. DOI:10.1016/j.neuron.2008.10.032. [GP]
- Fu, C. H., Vythelingum, G. N., Brammer, M. J., Williams, S. C., Amaro Jr., E., Andrew, C. M., Yaguez, L., van Haren, N. E., Matsumoto, K. & McGuire, P. K. (2006) An fMRI study of verbal self-monitoring: Neural correlates of auditory verbal feedback. *Cerebral Cortex* 16:969–77. [MIM]
- Gallese, V. (2008) Mirror neurons and the social nature of language: The neural exploitation hypothesis. *Social Neuroscience* 3:317–33. [aMJP]
- Gallese, V. (2009) Motor abstraction: A neuroscientific account of how action goals and intentions are mapped and understood. *Psychological Research* 73:486–98. [GD]
- Gallese, V. & Lakoff, G. (2005) The brain's concepts: The role of the sensory-motor system in reason and language. *Cognitive Neuropsychology* 22:455–79. [GD]
- Garrett, M. (1980) Levels of processing in speech production. In: *Language production, vol. 1. Speech and talk*, ed. B. Butterworth, pp. 177–220. Academic Press. [aMJP]
- Garrett, M. (2000) Remarks on the architecture of language production systems. In: *Language and the brain: Representation and processing*, ed. Y. Grodzinsky & L. P. Shapiro, pp. 31–69. Academic Press. [aMJP]
- Garrett, M. F. (1975) The analysis of sentence production. In: *The psychology of learning and motivation*, ed. G. H. Bower, pp. 133–75. Academic Press. [GSD]
- Garrod, S. & Anderson, A. (1987) Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition* 27:181–218. [aMJP]
- Garrod, S. & Pickering, M. J. (2004) Why is conversation so easy? *Trends in Cognitive Sciences* 8(1):8–11. [GP, aMJP]
- Garrod, S. & Pickering, M. J. (2009) Joint action, interactive alignment and dialogue. *Topics in Cognitive Science* 1:292–304. [KJA, aMJP]
- Gaskell, G. (2007) *Oxford handbook of psycholinguistics*. Oxford University Press. [aMJP]
- Gehring, W. J., Goss, B., Coles, M. G. H., Meyer, D. E. & Donchin, E. (1993) A neural system for error detection and compensation. *Psychological Science* 4:385–90. [F-XA]
- Gergely, G., Bekkering, H. & Király, I. (2002) Developmental psychology: Rational imitation in preverbal infants. *Nature* 415:755. [aMJP]
- Gerhardt, S. (2004) *Why love matters: How affection shapes a baby's brain*. Brunner-Routledge. [KJA]
- Gertner, Y. & Fisher, C. (2012) Predicted errors in children's early sentence comprehension. *Cognition* 124:85–94. [SMM]
- Geschwind, D. H. & Levitt, P. (2007) Autism spectrum disorders: Developmental disconnection syndromes. *Current Opinion in Neurobiology* 17:103–11. [KJA]
- Geva, S., Bennett, S., Warburton, E. a. & Patterson, K. (2011) Discrepancy between inner and overt speech: Implications for post-stroke aphasia and normal language processing. *Aphasiology* 25(3):323–43. DOI:10.1080/02687038.2010.511236. [GMO]
- Gibbon, F. E. (1999) Undifferentiated lingual gestures in children with articulation/phonological disorders. *Journal of Speech, Language and Hearing Research* 42(2):382. [SK]
- Gibson, E. (1998) Linguistic complexity: Locality of syntactic dependencies. *Cognition* 68:1–76. [aMJP]

- Gibson, E. & Hickok, G. (1993) Sentence processing with empty categories. *Language and Cognitive Processes* 8:147–61. [GH]
- Giles, H., Coupland, N. & Coupland, J. (1991) Accommodation theory: Communication, context, and consequence. In *Contexts of accommodation*, ed. H. Giles, J. Coupland & N. Coupland, pp. 1–68. Cambridge University Press. [YK]
- Glenberg, A. M. (2010) Embodiment as a unifying perspective for psychology. *Wiley Interdisciplinary Reviews: Cognitive Science* 1:586–96. [GD]
- Glenberg, A. M. (2011) How reading comprehension is embodied and why that matters. *International Electronic Journal of Elementary Education* 4:5–18. [ASD]
- Glenberg, A. M. & Gallese, V. (2012) Action-based language: A theory of language acquisition, comprehension, and production. *Cortex* 48(7):905–22. DOI:10.1016/j.cortex.2011.04.010. [ASD, arMJP]
- Glenberg, A. M. & Kaschak, M. P. (2002) Grounding language in action. *Psychonomic Bulletin & Review* 9:558–65. [aMJP]
- Goffman, L. (2010) Dynamic interaction of motor and language factors in normal and disordered development. In: *Speech motor control: New developments in basic and applied research*, ed. B. Maassen & P. van Lieshout, pp. 137–52. Oxford University Press. [SK]
- Goldberg, A. E. (1995) *Constructions: A construction grammar approach to argument structure*. University of Chicago Press. [MAJ, aMJP]
- Goldberg, A. E. (2006) *Constructions at work: The nature of generalization in language*. Oxford University Press. [MAJ]
- Goldberg, A. E. (2011) Corpus evidence of the viability of statistical preemption. *Cognitive Linguistics* 22:131–54. [MAJ]
- Goldberg, A. E., Casenhiser, D. M. & Sethuraman, N. (2005) The role of prediction in construction-learning. *Journal of Child Language* 32:407–26. [MAJ]
- Goldin-Meadow, S. (2005) What language creation in the manual modality tells us about the foundations of language. *The Linguistic Review* 22:199–226. [GD]
- Goldman, A. I. (2006) *Simulating minds. The philosophy, psychology, and neuroscience of mindreading*. Oxford University Press. [aMJP]
- Gollan, T. H., Montoya, R. I., Fennema-Notestine, C. & Morris, S.K. (2005) Bilingualism affects picture naming but not picture classification. *Memory & Cognition* 33:1220–34. [JF]
- Gollan, T. H., Montoya, R. I. & Werner, G. A. (2002) Semantic and letter fluency in Spanish-English bilinguals. *Neuropsychology* 16:562–76. [JF]
- Gomez, R. L. & Gerken, L. (1999) Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition* 70:109–35. [MAJ]
- Goodwin, C. (1979) The interactive construction of a sentence in natural conversation. In: *Everyday language: Studies in ethnomethodology*, ed. G. Psathas, 97–121. Irvington Publishers. [CH]
- Gordon, R. (1986) Folk psychology as simulation. *Mind and Language* 1:158–71. [aMJP]
- Graf Estes, K., Evans, J. L., Alibali, M. W. & Saffran, J. R. (2007) Can infants map meaning to newly segmented words? *Psychological Science* 18:254–60. [MAJ]
- Graf, M., Schütz-Bosbach, S. & Prinz, W. (2010) Motor involvement in object perception: Similarity and complementarity. In: *Grounding sociality: Neurons, minds, and culture*, ed. G. Semin & G. Echterhoff, pp. 27–52. Psychology Press. [aMJP]
- Grant, E. R. & Spivey, M. J. (2003) Eye movements and problem solving: Guiding attention guides thought. *Psychological Science* 14:462–66. [HK]
- Green, A., Straube, B., Weis, S., Jansen, A., Willmes, K., Konrad, K. & Kircher, T. (2009) Neural integration of iconic and unrelated coverbal gestures: A functional MRI study. *Human Brain Mapping* 30:3309–24. [ASD]
- Green, D. W. (1986) Control, activation and resource: A framework and a model for the control of speech in bilinguals *Brain and Language* 27:210–23. [JF]
- Green, J. R., Moore, C. A., Higashikawa, M. & Steeve, R. W. (2000) The physiologic development of speech motor control: Lip and jaw coordination. *Journal of Speech, Language, and Hearing Research* 43(1):239. [SK]
- Green, J. R. & Nip, I. B. (2010) Some organizational principles in early speech development. In: *Speech Motor Control: New developments in basic and applied research*, ed. B. Maassen & P. van Lieshout, pp. 172–88. Oxford University Press. [SK]
- Greene, S. B., McKoon, G. & Ratcliff, R. (1992) Pronoun resolution and discourse models. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 18:266–83. [SOY]
- Gregoromichelaki, E., Kempson, R., Purver, M., Mills, J. G., Cann, R., Meyer-Viol, W. & Healey, P. C. T. (2011) Incrementality and intention-recognition in utterance processing. *Dialogue and Discourse* 2:199–233. [aMJP]
- Grice, H. P. (1975) Logic and conversation. In: *Syntax and semantics 3: Speech acts*, ed. P. Cole & J. L. Morgan, pp. 41–58. Academic Press. [GE]
- Griffin, Z. M. & Bock, K. (2000) What the eyes say about speaking. *Psychological Science* 11:274–79. [HK]
- Grimm, A., Müller, A., Hamann, C. & Ruigendijk, E. (2011) *Production-comprehension asymmetries in child language*. De Gruyter. [SMM]
- Grossberg, S. (1980) How does a brain build a cognitive code? *Psychological Review* 87(1):1–51. [JB, rMJP]
- Grush, R. (2004) The emulation theory of representation: Motor control, imagery, and perception. *Behavioral and Brain Sciences* 27(3):377–96. [GP, aMJP]
- Guenther, F. H., Hampson, M. & Johnson, D. (1998) A theoretical investigation of reference frames for the planning of speech movements. *Psychological Review* 105(4):611. [TF]
- Guo T. & Peng D. (2006) ERP evidence for parallel activation of two languages in bilingual speech production. *NeuroReport* 17:1757–60. [JF]
- Häberle, A., Schütz-Bosbach, S., Laboisière, R. & Prinz, W. (2008) Ideomotor action in cooperative and competitive settings. *Social Neuroscience* 3:26–36. [aMJP]
- Hale, J. (2001) A probabilistic early parser as a psycholinguistic model. In: *North American Chapter of the Association for Computational Linguistics (NAACL)*, Vol. 2, pp. 1–8. Association for Computational Linguistics. [TF]
- Hale, J. (2006) Uncertainty about the rest of the sentence. *Cognitive Science* 30:609–42. [aMJP]
- Halsband, U., Schmitt, J., Weyers, M., Binkofski, F., Grutzner, G. & Freund, H. J. (2001) Recognition and imitation of pantomimed motor acts after unilateral parietal and premotor lesions: A perspective on apraxia. *Neuropsychologia* 39(2):200–16. [MP]
- Hamilton, A. C. & Martin, R. C. (2007) Proactive interference in a semantic short-term memory deficit: Role of semantic and phonological relatedness. *Cortex* 43:112–23. [LRS]
- Hanna, J. E. & Brennan, S. E. (2007) Speakers' eye gaze disambiguates referring expressions early during face-to-face conversation. *Journal of Memory and Language* 57:596–615. [HK]
- Hanna, J. E., Tanenhaus, M. K. & Trueswell, J. C. (2003) The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language* 49:43–61. [aMJP, SOY]
- Harley, T. (2008) *The psychology of language: From data to theory*, 3rd Edn., Psychology Press. [GSD, aMJP]
- Hartsuiker, R. J. & Kolk, H. H. J. (2001) Error monitoring in speech production: A computational test of the Perceptual Loop Theory. *Cognitive Psychology* 42:11357. [RJH, aMJP]
- Hartsuiker, R. J. & Notebaert, L. (2010) Lexical access problems lead to disfluencies in speech. *Experimental Psychology* 57:169–77. [RJH]
- Haruno, M., Wolpert, D. M. & Kawato, M. (2001) MOSAIC Model for sensorimotor learning and control. *Neural Computation* 13:2201–20. [aMJP]
- Haruno, M., Wolpert, D. M. & Kawato, M. (2003) Hierarchical MOSAIC for movement generation. *International Congress Series* 1250:575–90. [F-XA, aMJP]
- Hashimoto, Y. & Sakai, K. L. (2003) Brain activations during conscious self-monitoring of speech production with delayed auditory feedback: An fMRI study. *Human Brain Mapping* 20:22–28. [MIM]
- Hasson, U., Skipper, J. I., Nusbaum, H. C. & Small, S. L. (2007) Abstract coding of audiovisual speech: Beyond sensory representation. *Neuron* 56:1116–26. [ASD]
- Haueisen, J. & Knösche, T. R. (2001) Involuntary motor activity in pianists evoked by music perception. *Journal of Cognitive Neuroscience* 13:786–92. [aMJP]
- Hawkins, J. A. (1994) *A performance theory of order and constituency*. Cambridge University Press. [aMJP]
- Hay, J. F., Pelucchi, B., Graf Estes, K. & Saffran, J. R. (2011) Linking sounds to meaning: Infant statistical learning in a natural language. *Cognitive Psychology* 63:93–106. [MAJ]
- Healey, P. C. T., Purver, M. & Howes, C. (2010) Structural divergence in dialogue. *Proceedings of 20th Annual Meeting of the Society for Text & Discourse*. [CH]
- Heim, S., Opitz, B., Müller, K. & Friederici, A. D. (2003) Phonological processing during language production: fMRI evidence for a shared production-comprehension network. *Cognitive Brain Research* 12:285–29. [aMJP]
- Heinks-Maldonado, T. H., Nagarajan, S. S. & Houde, J. F. (2006) Magnetoencephalographic evidence for a precise forward model in speech production. *NeuroReport* 17(13):1375–79. [JPdR, RJH, MIM, arMJP, KS]
- Heller, D., Grodner, D. & Tanenhaus, M. K. (2008) The role of perspective in identifying domains of references. *Cognition* 108:831–36. [SOY]
- Hepper, P. G., Scott, D. & Shahidullah, S. (1993) Newborn and fetal response to maternal voice. *Journal of Reproductive and Infant Psychology* 11:147–53. [KJA]
- Heyes, C. (2010) Where do mirror neurons come from? *Neuroscience & Biobehavioral Reviews* 34(4):575–83. [GH]
- Hickok, G. (2009a) Eight problems for the mirror neuron theory of action understanding in monkeys and humans. *Journal of Cognitive Neuroscience* 21(7):1229–43. DOI: 10.1162/jocn.2009.21189. [ASD, CH, MP]
- Hickok, G. (2009b) The functional neuroanatomy of language. *Physics of Life Reviews* 6(3):121–43. [ASD]
- Hickok, G. (2012a) Computational neuroanatomy of speech production. *Nature Reviews Neuroscience* 13(2):135–45. [GH]
- Hickok, G. (2012b) The cortical organization of speech processing: Feedback control and predictive coding the context of a dual-stream model. *Journal of Communication Disorders* 45:393–402. [GH]

- Hickok, G., Houde, J. & Rong, F. (2011) Sensorimotor integration in speech processing: Computational basis and neural organization. *Neuron* 69(3):407–22. DOI:10.1016/j.neuron.2011.01.019. [ASD, GH]
- Hickok, G. & Poeppel, D. (2000) Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences* 4(4):131–38. [ASD]
- Hickok, G. & Poeppel, D. (2004) Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition* 92 (1–2):67–99. [ASD]
- Hickok, G. & Poeppel, D. (2007) The cortical organization of speech processing. *Nature Reviews Neuroscience* 8(5):393–402. [ASD, GH]
- Higgins, E. T. (1981) The “communication game”: Implications for social cognition and persuasion. In: *Social cognition: The Ontario Symposium*, Vol. 1, ed. E. T. Higgins, C. P. Herman & M. P. Zanna, pp. 343–92. Erlbaum. [GE]
- Hirano, S., Kojima, H., Naito, Y., Honjo, I., Kamoto, Y., Okazawa, H., Ishizu, K., Yonekura, Y., Nagahama, Y., Fukuyama, H. & Konishi, J. (1997) Cortical processing mechanism for vocalization with auditory verbal feedback. *NeuroReport* 8(9–10):2379–82. [MIM]
- Hockett, C. F. (1967) Where the tongue slips, there slip I. In: *To honor Roman Jakobson*, pp. 910–36. Mouton. [GMO]
- Holle, H., Gunter, T. C., Rüschmeyer, S. A., Hennenlotter, A. & Iacoboni, M. (2008) Neural correlates of the processing of co-speech gestures. *NeuroImage* 39(4):2010–24. [ASD]
- Holtgraves, T. & Kashima, Y. (2008) Language, meaning and social cognition. *Personality and Social Psychology Review* 12:73–94. [YK]
- Holtgraves, T. M. (2002) *Language as social action: Social psychology and language use*. Erlbaum. [GE]
- Hommel, B., Müsseler, J., Aschersleben, G. & Prinz, W. (2001) The theory of event coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences* 24:849–78. [aMJP]
- Horton, W. S. & Gerrig, R. J. (2005) The impact of memory demands on audience design during language production. *Cognition* 96:127–42. [LRS]
- Hostetter, A. B. & Alibali, M. W. (2008) Visual embodiment: Gesture as simulated action. *Psychonomic Bulletin & Review* 15:495–514. [aMJP]
- Hough, J. (2011) Incremental semantics driven natural language generation with self-repairing capability. *Proceedings of RANLP 2011 Student Conference*, September 2011, Hissar, Bulgaria, 79–84. [CH]
- Howes, C., Healey, P. G. T., Purver, M. & Eshghi, A. (2012) Finishing each other's... Responding to incomplete contributions in dialogue. In: *Proceedings of the 34th Annual Conference of the Cognitive Science Society*, August 2012, Sapporo, Japan, 479–85. [CH]
- Howes, C., Purver, M., Healey, P. G. T., Mills, G. J. & Gregoromichelaki, E. (2011) Incrementality in dialogue: Evidence from compound contributions. *Dialogue and Discourse* 2:279–311. [aMJP]
- Huetting, F. & Hartsuiker, R. J. (2010) Listening to yourself is like listening to others: External, but not internal, verbal self-monitoring is based on speech perception. *Language and Cognitive Processes* 25:347–74. [RJH, arMJP]
- Huetting, F. & Janse, E. (2012) Anticipatory eye movements are modulated by working memory capacity: Evidence from older adults. Paper presented at the *18th Architectures and Mechanisms for Language Processing Conference*, Riva del Garda, Italy. [NM]
- Huetting, F. & McQueen, J. M. (2007) The tug of war between phonological, semantic and shape information in language-mediated visual search. *Journal of Memory and Language* 57:460–82. [RJH, NM, aMJP]
- Huetting, F., Rommers, J. & Meyer, A. S. (2011) Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica* 137:151–71. [HK, aMJP]
- Hurley, S. (2008a) The shared circuits model (SCM): How control, mirroring, and simulation can enable imitation, deliberation, and mindreading. *Behavioral and Brain Sciences* 31(01):1–22. [GD, YK, arMJP, HR]
- Hurley, S. (2008b) Understanding simulation. *Philosophy and Phenomenological Research* 77:755–74. [aMJP]
- Iacoboni, M. (2008) The role of premotor cortex in speech perception: Evidence from fmri and rtms. *Journal of Physiology (Paris)*, 102:31–34. [ASD]
- Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C. & Rizzolatti, G. (2005) Grasping the intentions of others with one's own mirror neuron system. *PLoS Biology* 3(3):e79. DOI: 10.1371/journal.pbio.0030079. [MP]
- Indefrey P. & Levelt, W. J. M. (2004) The spatial and temporal signatures of word production components. *Cognition* 92:101–44. [aMJP, KS]
- Ito, T., Tiede, M. & Ostry, D. J. (2009) Somatosensory function in speech perception. *Proceedings of the National Academy of Sciences* 106:1245–48. [aMJP]
- Iverson, J. (2010) Developing language in a developing body: The relationship between motor development and language development. *Journal of Child Language* 37(2):229–61. [SK]
- Iverson, J. & Braddock, B. A. (2011) Gesture and motor skill in relation to language in children with language impairment. *Journal of Speech, Language, and Hearing Research* 54(1):72–86. [SK]
- Iverson, J. M. & Thelen, E. (1999) Hand, mouth and brain. The dynamic emergence of speech and gesture. *Journal of Consciousness Studies* 6:11(12):19–40. [SK]
- Jackendoff, R. (2002) *Foundations of language: Brain, meaning, grammar, evolution*. Oxford University Press. [GD]
- Jackendoff, R. (2007) Linguistics in cognitive science: The state of the art. *The Linguistic Review* 24:347–401. [GD]
- Jaeger, T. F. (2010) Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology* 61:23–62. [TFJ, aMJP]
- Jaeger, T. F. & Snider, N. (2013) Alignment as a consequence of expectation adaptation: Syntactic priming is affected by the prime's prediction error given both prior and recent experience. *Cognition* 127(1):57–83. [TFJ]
- Janssen N. & Barber, H. A. (2012) Phrase frequency effects in language production. *PLoS ONE* 7:e33202. [SMM]
- Jonides, J. & Nee, D. E. (2006) Brain mechanisms of proactive interference in working memory. *Neuroscience* 139(1):181–93. [LRS]
- Jordan, M. I. & Rumelhart, D. E. (1992) Forward models: Supervised learning with a distal teacher. *Cognitive Science*, 16: 307–54. [FC, rMJP]
- Kaiser, E. & Trueswell, J. C. (2004) The role of discourse context in the processing of a flexible word-order language. *Cognition* 94:113–47. [aMJP]
- Kamide, Y., Altmann, G. T. M. & Haywood, S. L. (2003) Prediction and thematic information in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language* 49:133–56. [HK, aMJP]
- Kamide, Y., Scheepers, C. & Altmann, G. T. M. (2003) Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from German and English. *Journal of Psycholinguistic Research* 32(1):37–55. [FC]
- Kan, I. P. & Thompson-Schill, S. L. (2004) Effect of name agreement on prefrontal activity during overt and covert picture naming. *Cognitive, Affective & Behavioral Neuroscience* 4(1):43–57. [GMO]
- Kashima, Y. & Lan, Y. (in press) Communication and language use in social cognition. In: *The Oxford handbook of social cognition*, ed. D. Carlson. Oxford University Press. [YK]
- Kawato, M. (1999) Internal models for motor control and trajectory planning. *Current Opinion in Neurobiology* 9(6):718–27. [GH]
- Kawato, M., Furawaka, K. & Suzuki, R. (1987) A hierarchical neural network model for the control and learning of voluntary movements. *Biological Cybernetics* 56: 1–17. [rMJP]
- Kawato, M., Maeda, Y., Uno, Y. & Suzuki, R. (1990) Trajectory formation of arm movement by cascade neural network model based on minimum torque-change criterion. *Biological Cybernetics*, 62: 275–88. [rMJP]
- Kemmerer, D. (2010) How words capture visual experience: The perspective from cognitive neuroscience. In: *Words and the world: How words capture human experience*, ed. B. Malt & P. Wolff, pp. 289–329. Oxford University Press. [GD]
- Kempen, G. & Huijbers, P. (1983) The lexicalization process in sentence production and naming: Indirect election of words. *Cognition* 14:185–209. [GSD]
- Keysar, B., Barr, D. J., Balin, J. A. & Brauner, J. S. (2000) Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science* 11:32–38. [aMJP]
- Kidd, E. (2012) Implicit statistical learning is directly associated with the acquisition of syntax. *Developmental Psychology* 48(1):171–84. [FC]
- Kilner, J. M. (2011) More than one pathway to action understanding. *Trends in Cognitive Sciences* 15(8):352–57. DOI: 10.1016/j.tics.2011.06.005. [MP]
- Kilner, J. M., Paulignan, Y. & Blakemore, S.-J. (2003) An interference effect of observed biological movement on action. *Current Biology* 13:522–25. [aMJP]
- Kim, A. & Lai, V. (2012) Rapid interactions between lexical semantic and word form analysis during word recognition in context: Evidence from ERPs. *Journal of Cognitive Neuroscience* 24: 1104–12. [aMJP]
- Kiparsky, P. (1982) Lexical morphology and phonology. In: *Linguistics in the Morning Calm*, ed. I.-S. Yang, pp. 3–91. Hanshin. [MAJ]
- Kleinschmidt, D., Fine, A. B. & Jaeger, T. F. (2012) A belief-updating model of adaptation and cue combination in syntactic comprehension. *Proceedings of the 34rd Annual Meeting of the Cognitive Science Society (CogSci2012)*, 605–10. [TFJ]
- Kleinschmidt, D. & Jaeger, T. F. (2011) A Bayesian belief updating model of phonetic recalibration and selective adaptation. *Proceedings of the Cognitive Modeling and Computational Linguistics Workshop at ACL, Portland, OR, June 23rd*, 10–19. [TFJ]
- Knoblich, G. & Flach, R. (2001) Predicting action effects: Interaction between perception and action. *Psychological Science* 12:467–72. [aMJP]
- Knoblich, G., Öllinger, M. & Spivey, M. J. (2005) Tracking the eyes to obtain insight into insight problem solving. In: *Cognitive processes in eye guidance*, ed. G. Underwood, pp. 355–75. Oxford University Press. [HK]
- Knoblich, G., Seigerschmidt, E., Flach, R. & Prinz, W. (2002) Authorship effects in the prediction of handwriting strokes: Evidence for action simulation during action perception. *Quarterly Journal of Experimental Psychology* 55A:1027–46. [aMJP]

- Knoeferle, P. & Crocker, M. W. (2006) The coordinated interplay of scene, utterance, and world knowledge: Evidence from eye tracking. *Cognitive Science* 30:481–529. [HK]
- Knoeferle, P., Crocker, M. W., Scheepers, C. & Pickering, M. J. (2005) The influence of the immediate visual context on incremental thematic role-assignment: Evidence from eye-movements in depicted events. *Cognition* 95:95–127. [aMJJP]
- Knoeferle, P. & Kreysa, H. (2012) Can speaker gaze modulate syntactic structuring and thematic role assignment during spoken sentence comprehension? *Frontiers in Psychology* 3:538. [HK]
- Krishnan, S., Alcock, K. J., Mercure, E., Leech, R., Barker, E., Karmiloff-Smith, A. & Dick, F. (in press) Articulating novel words: Oromotor contributions to individual differences in nonword repetition. *Journal of Speech, Language and Hearing Research*. [SK]
- Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., Stolyarova, E. I., Sundberg, U. & Lacerda, F. (1997) Cross-language analysis of phonetic units in language addressed to infants. *Science* 277(5326):684–86. [GP]
- Kutas, M., DeLong, K. A. & Smith, N. J. (2011). A look around at what lies ahead: Prediction and predictability in language processing. In: *Predictions in the brain: Using our past to generate a future*, ed. M. Bar, pp. 190–207. Oxford University Press. [aMJJP]
- Lakin, J. & Chartrand, T. L. (2003) Using nonconscious behavioral mimicry to create affiliation and rapport. *Psychological Science* 14:334–39. [aMJJP]
- Langton, S. R. H., Watt, R. J. & Bruce, V. (2000) Do the eyes have it? Cues to the direction of social attention. *Trends in Cognitive Sciences* 4:50–59. [HK]
- Lau, E., Stroud, C., Plesch, S. & Phillips, C. (2006) The role of structural prediction in rapid syntactic analysis. *Brain and Language* 98:74–88. [aMJJP]
- Laver, J. D. M. (1980) Monitoring systems in the neurolinguistic control of speech production. In: *Errors in linguistic performance: Slips of the tongue, ear, pen, and hand*, ed. V. A. Fromkin, Academic Press. [aMJJP]
- Lebeltel, O., Bessière, P., Diard, J. & Mazer, E. (2004) Bayesian robot programming. *Autonomous Robots* 16(1):49–79. [RL]
- Lerner, G. (1991) On the syntax of sentences-in-progress. *Language in Society* 20(3):441–58. [CH]
- Leroi-Gourhan, A. (1964) *Le Geste et la Parole I Technique et Langage*, Paris, Albin Michel (coll. Sciences d'aujourd'hui), p. 323. [MP]
- Leroi-Gourhan, A. (1965) *Le Geste et la Parole II La Memoire et les rythmes*, Paris, Albin Michel (coll. Sciences d'aujourd'hui), p. 285. [MP]
- Lesage, E., Morgan, B. E., Olson, A. C., Meyer, A. S. & Miall, R. C. (2012) Cerebellar rTMS disrupts predictive language processing. *Current Biology* 22, R794–95. [rMJJP]
- Levelt, W. J. M. (1983) Monitoring and self-repair in speech. *Cognition* 14:41–104. [MIM, GMO, aMJJP]
- Levelt, W. J. M. (1989) *Speaking: From intention to articulation*. MIT Press. [JPdR, RJH, arMJJP]
- Levelt, W. J. M., Roelofs, A. & Meyer, A. S. (1999) A theory of lexical access in speech production. *Behavioral and Brain Sciences* 22(1):1–75. [GSD, GMO, aMJJP, KS]
- Levinson, S. C. (1983) *Pragmatics*. Cambridge University Press. [JPdR]
- Levinson, S. C. (1995) Interaction biases in human thinking. In: *Social intelligence and interaction*, ed. E. N. Goody, pp. 221–60. Cambridge University Press. [JPdR]
- Levinson, S. C. (2006) On the human “interaction engine.” In: *Roots of human sociality: Culture, cognition and interaction*, ed. N. J. Enfield & S. C. Levinson (Cur.), pp. 39–69. Berg. [HK, GP]
- Levy, R. (2008) Expectation-based syntactic comprehension. *Cognition* 106(3):1126–77. [TFJ, aMJJP, HR]
- Lew-Williams, C. & Fernald, A. (2007) Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science* 18(3):193–98. [FC]
- Lewis, J. D. & Elman, J. L. (2001) Learnability and the statistical structure of language: Poverty of stimulus arguments revisited. *Proceedings of the 26th Annual Conference on Language Development*. [MAJ]
- Lichtheim, L. (1885) On aphasia. *Brain* 7(4):433–54. [HR]
- Lindblom, B. (1990) Explaining phonetic variation: A sketch of the H&H theory. *Speech production and speech modelling* 55:40339. [TFJ]
- MacDonald, M. C., Pearlmutter, N. J. & Seidenberg, M. S. (1994) The lexical nature of syntactic ambiguity resolution. *Psychological Review* 101:676–703. [aMJJP]
- MacGregor, L. J., Pulvermuller, F., van Casteren, M. & Shtyrov, Y. (2012) Ultra-rapid access to words in the brain. *Nature Communications* 3:711. [KS]
- MacKay, D. G. (1981) The problem of rehearsal or mental practice. *Journal of Motor Behavior* 13(4):274–85. [GMO]
- MacKay, D. G. (1982) The problems of flexibility, fluency, and speed-accuracy trade-off in skilled behaviors. *Psychological Review* 89:483–506. [aMJJP]
- MacKay, D. G. (1992) Constraints on theories of inner speech. In: *Auditory imagery*, ed. D. Reisberg, pp. 121–49. Erlbaum. [GMO]
- Magyar, L. & De Ruiter, J. P. (2012) Prediction of turn-ends based on anticipation of upcoming words. *Frontiers in Psychology* 3:376. [JPdR]
- Mahon, B. Z. & Caramazza, A. (2008) A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology-Paris* 102(1–3):59–70. DOI: 10.1016/j.jphysparis.2008.03.004. [MP]
- Mahon, B. Z. & Caramazza, A. (2009) Concepts and categories: A cognitive neuropsychological perspective. *Annual Review of Psychology* 60:27–51. DOI:10.1146/annurev.psych.60.110707.163532. [ASD]
- Mampe, B., Friederici, A. D., Christophe, A. & Wermke, K. (2009) Newborns' cry melody is shaped by their native language. *Current Biology* 19:1994–97. [KJA]
- Mani, N., Durrant, S. & Floccia, C. (2012) Activation of phonological and semantic codes in toddlers. *Journal of Memory and Language* 66:612–22. [NM]
- Mani, N. & Huettig, F. (2012) Prediction during language processing is a piece of cake – but only for skilled producers. *Journal of Experimental Psychology: Human Perception and Performance* 38: 843–47. [FC, NM, SMM, rMJJP]
- Mani, N. & Plunkett, K. (2010) In the infant's mind's ear: Evidence for implicit naming in infancy. *Psychological Science* 21:908–13. [NM]
- Mar, R. A. (2004) The neuropsychology of narrative: Story comprehension, story production and their interrelation. *Neuropsychologia* 42:1414–34. [aMJJP]
- Marchman, V. A. & Fernald, A. (2008) Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental Science* 11(3):F9–16. [FC]
- Marcotte, J. (2005) Causative alternation errors as event-driven construction paradigm completions. Stanford, Ph.D. dissertation. [MAJ]
- Marshall, R. C., Rappaport, B. Z. & Garcia-Bunuel, L. (1985) Self-monitoring behavior in a case of severe auditory agnosia with aphasia. *Brain and Language* 24:297–313. [RJH]
- Marslen-Wilson, W. D. (1973) Linguistic structure and speech shadowing at very short latencies. *Nature* 244:522–23. [JPdR]
- Marslen-Wilson, W. D. & Welsh, A. (1978) Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology* 10:29–63. [aMJJP]
- Martin, H. F. & Zwaan, R. A. (2008) Embodied language: A review of the role of the motor system in language comprehension. *The Quarterly Journal of Experimental Psychology* 61:825–50. [GD]
- Masataka, N. (2001) Why early linguistic milestones are delayed in children with Williams syndrome: Late onset of hand banging as a possible rate-limiting constraint on the emergence of canonical babbling. *Developmental Science* 4(2):158–164. [SK]
- Mattson, M. & Baars, B. J. (1992) Error-minimizing mechanisms: Boosting or editing? In *Experimental slips and human error: Exploring the architecture of volition*, ed. B. J. Baars, pp. 263–87. Plenum. [RJH]
- McCauley, S. M. & 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*, ed. L. Carlson, C. Hölscher & T. Shipley, pp. 1619–24. Cognitive Science Society. [SMM]
- McCauley, S. M. & Christiansen, M. H. (submitted) *Language learning as language use: A computational model of children's comprehension and production of language*. Manuscript in preparation, Cornell University. [SMM]
- McDonald, S. A. & Shillcock, R. C. (2003) Eye movements reveal the on-line computation of lexical probabilities. *Psychological Science* 14:648–52. [NM]
- McGuire, P. K., Silbersweig, D. A. & Frith, C. D. (1996) Functional neuroanatomy of verbal self-monitoring. *Brain* 119(Pt. 3):907–17. [MIM]
- McHale, J., Fivaz-Depeursinge, E., Dickstein S., Robertson, J. & Daley, M. (2008) New evidence for the social embeddedness of infants' early triangular capacities. *Family Process* 47:445–63. [KJA]
- 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. [AMT]
- Meister, I. G., Wilson, S. M., Deblieck, C., Wu, A. D. & Iacoboni, M. (2007) The essential role of premotor cortex in speech perception. *Current Biology* 17:1692–96. [ASD, RL]
- Meltzoff, A. N. & Moore, M. K. (1977) Imitation of facial and manual gestures by human neonates. *Science* 198:75–8. [KJA]
- Melzer, A., Prinz, W. & Daum, M. M. (2012) Production and perception of contralateral reaching: A close link by 12 months of age. *Infant Behavior and Development* 35:570–79. [NM]
- Menenti, L., Gierhan, S. M. E., Segaert, K. & Hagoort, P. (2011) Shared language: Overlap and segregation of the neuronal infrastructure for speaking and listening revealed by fMRI. *Psychological Science* 22:1173–82. [aMJJP]
- Meringer, R. & Meyer, K. (1995) *Versprechen und verstehen*. Behrs Verlag. [GMO]
- Meteyard, L., Cuadrado, S. R., Bahrami, B. & Vigliocco, G. (2012) Coming of age: A review of embodiment and the neuroscience of semantics. *Cortex* 48:788–804. [GD]

- Metzing, C. & Brennan, S. E. (2003) When conceptual pacts are broken: Partner-specific effects in the comprehension of referring expressions. *Journal of Memory and Language* 49:201–13. [aMJP]
- Meuter, R. & Allport, A. (1999) Bilingual language switching in naming: Asymmetric costs of language selection. *Journal of Memory and Language* 40:25–40. [JF]
- Meyer, A. S., Sleiderink, A. M. & Levelt, W. J. M. (1998) Viewing and naming objects: Eye movements during noun phrase production. *Cognition* 66: B25–33. [HK]
- Miall, R. C., Stanley, J., Todhunter, S., Levick, C., Lindo, S. & Miall, J. D. (2006) Performing hand actions assists the visual discrimination of similar hand postures. *Neuropsychologia* 44:966–76. [aMJP]
- Miall, R. C. & Wolpert, D. M. (1996) Forward models for physiological motor control. *Neural Networks* 9: 1265–79. [rMJP]
- Milner, A. D. & Goodale, M. A. (1995) *The visual brain in action*. Oxford University Press. [GH]
- Mirman, D., Magnuson, J., Graf Estes, K. & Dixon, J. A. (2008) The link between statistical segmentation and word learning in adults. *Cognition* 108:271–80. [MAJ]
- Mishra, R. K., Singh, N., Pandey, A. & Huettig, F. (2012) Spoken language-mediated anticipatory eye movements are modulated by reading ability: Evidence from Indian low and high literates. *Journal of Eye Movement Research* 5(1):1–10. [NM]
- Misyak, J. B., Christiansen, M. H. & Tomblin, J. B. (2010) Sequential expectations: The role of prediction-based learning in language. *Topics in Cognitive Science* 2:138–53. [MAJ]
- Mitterer, H. & Ernestus, M. (2008) The link between speech perception and production is phonological and abstract: Evidence from the shadowing task. *Cognition* 109:168–73. [AMT]
- Moore, R. K. (2007) PRESENCE: A human-inspired architecture for speech-based human-machine interaction. *IEEE Transactions on Computers* 56(9):1176–88. [GP]
- Motley, M. T., Camden, C. T. & Baars, B. J. (1982) Covert formulation and editing of anomalies in speech production: Evidence from experimentally elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior* 21:578–94. [aMJP]
- Mottonen, R. & Watkins, K. E. (2009) Motor representations of articulators contribute to categorical perception of speech sounds. *Journal of Neuroscience* 29(31):9819–25. DOI: 10.1523/JNEUROSCI.6018-08.2009. [MP, aMJP]
- Moulin-Frier, C., Laurent, R., Bessière, P., Schwartz, J.-L. & Diard, J. (2012) Adverse conditions improve distinguishability of auditory, motor and perceptuo-motor theories of speech perception: An exploratory Bayesian modeling study. *Language and Cognitive Processes* 27(7–8):1240–63. [RL]
- Mukamel, R., Ekstrom, A. D., Kaplan, J., Iacoboni, M. & Fried, I. (2010) Single-neuron responses in humans during execution and observation of actions. *Current Biology* 20:750–56. [aMJP]
- Nappa, R., Wessel, A., McEllood, K. L., Gleitman, L. R. & Trueswell, J. C. (2009) Use of speaker's gaze and syntax in verb learning. *Language Learning and Development* 5:203–34. [HK]
- Navab, A., Gillespie-Lynch, K., Johnson, S. P., Sigman, M. & Hutman, T. (2011) Eye-tracking as a measure of responsiveness to joint attention in infants at risk for autism. *Infancy* 17:416–31. [KJA]
- Neda, Z., Ravasz, Y., Brechet, T., Vicsek, T. & Barabasi, A. L. (2000) The sound of many hands clapping. *Nature* 403:849. [aMJP]
- Negri, G. A., Rumiati, R. I., Zadini, A., Ukmari, M., Mahon, B. Z. & Caramazza, A. (2007) What is the role of motor simulation in action and object recognition? Evidence from apraxia. *Cognitive Neuropsychology* 24(8):795–816. DOI: 10.1080/02643290701707412. [MP]
- Nelissen, N., Pazzaglia, M., Vandenbulcke, M., Sunaert, S., Fannes, K., Dupont, P., Aglioti, S. M. & Vandenberghe, R. (2010) Gesture discrimination in primary progressive aphasia: The intersection between gesture and language processing pathways. *Journal of Neuroscience* 30(18):6334–41. DOI: 10.1523/JNEUROSCI.0321-10.2010. [MP]
- Newman-Norlund, R. D., van Schie, H. T., van Zuijlen, A. M. J. & Bekkering, H. (2007) The mirror neuron system is more active during complementary compared with imitative action. *Nature Neuroscience* 10:817–18. [aMJP]
- Nielsen, K. (2011) Specificity and abstractness of VOT imitation. *Journal of Phonetics* 39:132–42. [AMT]
- Nieuwland, M. S., Otten, M. & Van Berkum, J. J. A. (2007) Who are you talking about? Tracking discourse-level referential processes with ERPs. *The Journal of Cognitive Neuroscience* 19:1–9. [SOY]
- Niv, Y. & Montague, P. R. (2008) Theoretical and empirical studies of learning. *Neuroeconomics: Decision making and the brain*, pp. 329–50. Elsevier. [MAJ]
- Nooteboom, S. G. (1969) The tongue slips into patterns. In: *Leyden studies in linguistics and phonetics*, ed. A. C. Sciarone, A. J. van Essen & A. A. van Raad, pp. 114–32. Mouton. [GMO]
- Novick, J. M., Kan, I. P., Trueswell, J. C. & Thompson-Schill, S. L. (2009) A case for conflict across multiple domains: memory and language impairments following damage to ventrolateral prefrontal cortex. *Cognitive Neuropsychology* 26(6):527–67. [LRS]
- Novick, J. M., Trueswell, J. C. & Thompson-Schill, S. L. (2005) Cognitive control and parsing: Reexamining the role of Broca's area in sentence comprehension. *Cognitive & Behavioral Neuroscience* 5(3):263–81. [LRS]
- Nozari, N. & Dell, G. S. (2009) More on lexical bias: How efficient can a “lexical editor” be? *Journal of Memory and Language* 60:291–307. [GSD]
- Nozari, N., Dell, G. S. & Schwartz, M. F. (2011) Is comprehension necessary for error detection? A conflict-based account of monitoring in speech production. *Cognitive Psychology* 63(1):1–33. DOI:10.1016/j.cogpsych.2011.05.001. [RJH, GMO, aMJP]
- O'Doherty, J. P., Dayan, P., Schultz, J., Deichmann, R., Friston, K. & Dolan, R. (2004) Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science* 304:452–54. [MAJ]
- Ondobaka, S. & Bekkering, H. (2012) Hierarchy of idea-guided action and perception-guided movement. *Frontiers in Psychology*. doi: 10.3389/fpsyg.2012.00579 [YK]
- Ondobaka, S., de Lange, F. P., Newman-Norlund, R. D., Wiemers, M. & Bekkering, H. (2011) Interplay between action and movement intentions during social interaction. *Psychological Science* 23: 30–35. [YK, rMJP]
- Oomen, C. C. E., Postma, A. & Kolk, H. H. J. (2005) Speech monitoring in aphasia: Error detection and repair behaviour in a patient with Broca's aphasia. In: *Phonological encoding and monitoring in normal and pathological speech*, ed. R. Hartsuiker, R. Bastiaanse, A. Postma & F. Wijnen, pp. 209–25. Psychology Press. [RJH]
- Oppenheim, G. M. (2012) The case for subphonemic attenuation in inner speech: Comment on Corley, Brocklehurst, and Moat (2011). *Journal of Experimental Psychology: Learning, Memory, and Cognition* 38(3):502–12. DOI:10.1037/a0025257. [GMO]
- Oppenheim, G. M. & Dell, G. S. (2008) Inner speech slips exhibit lexical bias, but not the phonemic similarity effect. *Cognition* 106(1):528–37. DOI:10.1016/j.cognition.2007.02.006. [GMO, aMJP]
- Oppenheim, G. M. & Dell, G. S. (2010) Motor movement matters: The flexible abstractness of inner speech. *Memory & cognition* 38(8):1147–60. DOI:10.1016/j.cognition.2007.02.006. [F-XA, GMO, aMJP]
- Pacherie, E. (2012) The phenomenology of joint action: Self-agency vs. joint-agency. In: *Joint attention: New developments*, ed. A. Seemann, pp. 343–89. MIT Press. [YK]
- Pagnoni, G., Zink, C. F., Montague, P. R. & Berns, G. S. (2002) Activity in human ventral striatum locked to errors of reward prediction. *Nature Neuroscience* 5:97–98. [MAJ]
- Panksepp, J. (2004) *Affective neuroscience: The foundations of human and animal emotions*. Oxford University Press. [KJA]
- Pardo, J. S. (2006) On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America* 119:2382–93. [aMJP]
- Parker Jones, Ö., Green, D. W., Grogan, A., Platsikas, C., Filippopolitis, K., Ali, N., Lee, H. L., Ramsden, S., Gazarian, K., Prejawa, S., Seghier, M. L. & Price, C. J. (2012) Where, when and why brain activation differs for bilinguals and monolinguals during picture naming and reading aloud. *Cerebral Cortex* 22:892–902. [JF]
- Paus, T., Perry, D. W., Zatorre, R. J., Worsley, K. J. & Evans, A. C. (1996) Modulation of cerebral blood flow in the human auditory cortex during speech: Role of motor-to-sensory discharges. *European Journal of Neuroscience* 8:2236–46. [aMJP]
- Pazzaglia, M., Pizzamiglio, L., Pes, E. & Aglioti, S. M. (2008b) The sound of actions in apraxia. *Current Biology* 18(22):1766–72. DOI: 10.1016/j.cub.2008.09.061. [MP]
- Pazzaglia, M., Smania, N., Corato, E. & Aglioti, S. M. (2008a) Neural underpinnings of gesture discrimination in patients with limb apraxia. *Journal of Neuroscience* 28(12):3030–41. DOI: 10.1523/JNEUROSCI.5748-07.2008. [MP]
- Pazzaglia, M. (2013) Action discrimination: Impact of apraxia. *Journal of Neurology, Neurosurgery & Psychiatry* 84(5):477–78. doi: 10.1136/jnnp-2012-304817 [MP]
- Pelucchi, B., Hay, J. F. & Saffran, J. R. (2009) Learning in reverse: Eight-month-old infants track backward transitional probabilities. *Cognition* 113:244–47. [SMM]
- Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Stockmann, E., Tiede, M., & Zandipour, M. (2004) The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts. *The Journal of the Acoustical Society of America* 116:2338. [TFJ]
- Peterson, R. R., Burgess, C., Dell, G. S. & Eberhard, K. A. (2001) Dissociation between syntactic and semantic processing during idiom comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 27:1223–37. [aMJP]
- Pezzulo, G. (2011a) Grounding procedural and declarative knowledge in sensori-motor anticipation. *Mind and Language* 26:78–114. [aMJP]
- Pezzulo, G. (2011b) The “interaction engine”: A common pragmatic competence across linguistic and non-linguistic interactions. *IEEE Transactions on Autonomous Mental Development* 4(2):105–23. [GP]
- Pezzulo, G. (2011c) Shared representations as coordination tools for interactions. *Review of Philosophy and Psychology* 2(2):303–33. [GP]

- Pezzulo, G., Barsalou, L. W., Cangelosi, A., Fischer, M. H., McRae, K. & Spivey, M. J. (2011) The mechanics of embodiment: A dialog on embodiment and computational modeling. *Frontiers in Psychology* 2(5):1–21. [GD]
- Pezzulo, G. & Dindo, H. (2011) What should I do next? Using shared representations to solve interaction problems. *Experimental Brain Research* 211(3):613–630. [GP]
- Pickering, M. J. & Garrod, S. (2004) Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences* 27(2):169–226. [CH, aMJJP, GP, SOY]
- Pickering, M. J. & Garrod, S. (2007) Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences* 11(3):105–110. [CH, aMJJP, JpDR]
- Pinker, S. (2007) *The stuff of thought: Language as a window into human nature*. Penguin. [GD]
- Plaut, D. C. & Kello, C. T. (1999) The emergence of phonology from the interplay of speech comprehension and production: A distributed connectionist approach. In: *The emergence of language*, ed. B. MacWhinney, pp. 381–415. Erlbaum. [FC, aMJJP]
- Poizner, H., Klima, E. S. & Bellugi, U. (1987) *What the hands reveal about the brain*. MIT Press. [GD]
- Postma, A. (2000) Detection of errors during speech production. A review of speech monitoring models. *Cognition* 77: 97–131. [aMJJP]
- Postma, A. & Noordanus, C. (1996) Production and detection of speech errors in silent, mouthed, noise-masked, and normal auditory feedback speech. *Language and Speech* 39(4):375–92. [GMO]
- Price, C. J. (2010) The anatomy of language: A review of 100 fMRI studies published in 2009. *Annals of the New York Academy of Sciences* 1191(1):62–88. [ASD]
- Price, C. J. (2012) A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *NeuroImage* 62(2):816–47. DOI: 10.1016/j.neuroimage.2012.04.062. [ASD]
- Prinz, J. (2012) Waiting for the self. In: *Consciousness and the self: New essays*, ed. Liu, J. L. and Perry, J. pp. 213–40. Cambridge University Press. [MIM]
- Prinz, W. (2006) What re enactment earns us. *Cortex* 42:515–18. [aMJJP]
- Przyrembel, M., Smallwood, J., Pauen, M. & Singer, T. (2012) Illuminating the dark matter of social neuroscience: Considering the problem of social interaction from philosophical, psychological and neuroscientific perspectives. *Frontiers in Human Neuroscience* 6:190. DOI:10.3389/fnhum.2012.00190. [KJA]
- Pulvermüller, F. (2005) Brain mechanisms linking language and action. *Nature Reviews Neuroscience*. 6(7):576–82. Review. [MP]
- Pulvermüller, F. & Fadiga, L. (2010) Active perception: Sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience* 11(5):351–60. DOI: 10.1038/nrn2811. [MP, aMJJP]
- Pulvermüller, F., Huss, M., Kherif, F., Moscoso del Prado Martin, F., Hauk, O. & Shtyrov, Y. (2006) Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences* 103(20):7865–70. [ASD, aMJJP]
- Pulvermüller, F. & Shtyrov, Y. (2006) Language outside the focus of attention: the mismatch negativity as a tool for studying higher cognitive processes. *Progress in Neurobiology* 79:49–71. [KS]
- Purver, M., Cann, R. & Kempson, R. (2006) Grammars as parsers: The dialogue challenge. *Research in Language and Computation* 4:289–326. [CH]
- Purver, M., Eshghi, A. & Hough, J. (2011) Incremental semantic construction in a dialogue system. *Proceedings of the 9th International Conference on Computational Semantics (IWCS)*. January 2011, Oxford, UK, 365–69. [CH]
- Ramscar, M., Yarlett, D., Dye, M., Denny, K. & Thorpe, K. (2010) The effects of feature-label-order and their implications for symbolic learning. *Cognitive Science* 34(6):909–57. [TFJ]
- Rapp, B. & Goldrick, M. (2000) Discreteness and interactivity in spoken word production. *Psychological Review* 107:460–99. [aMJJP]
- Rapp, B. & Goldrick, M. (2004) Feedback by any other name is still interactivity: A reply to Roelofs (2004). *Psychological Review* 111:573–78. [aMJJP]
- Rauschecker, A. M., Pringle, A. & Watkins, K. E. (2008) Changes in neural activity associated with learning to articulate novel auditory pseudowords by covert repetition. *Human brain mapping* 29(11):1231–42. DOI:10.1002/hbm.20460. [GMO]
- Rauschecker, J. P. (2011) An expanded role for the dorsal auditory pathway in sensorimotor control and integration. *Hearing Research* 271(1–2):16–25. DOI:10.1016/j.heares.2010.09.001. [ASD]
- Rauschecker, J. P. & Scott, S. K. (2009) Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience* 12(6):718–24. DOI:10.1038/nn.2331. [F-XA, ASD, GH]
- Rauschecker, J. P. & Tian, B. (2000) Mechanisms and streams for processing of “what” and “where” in auditory cortex. *Proceedings of the National Academy of Sciences* 97(22):11800. [ASD]
- Real, F. & Christiansen, M. H. (2007) Processing of relative clauses is made easier by frequency of occurrence. *Journal of Memory and Language*, 57:1–23. [SMM]
- Rescorla, R. A. & Wagner, A. R. (1972) A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In: *Classical conditioning II*, A. H. Black & F. Prokasy, pp. 64–99. Appleton-Century-Crofts. [MAJ]
- Rhode, H., Levy, R. & Kehler, A. (2011) Anticipating explanations in relative clause processing. *Cognition* 118:339–58. [LRS]
- Richardson, D. C. & Dale, R. (2005) Looking to understand: The coupling between speakers’ and listeners’ eye movements and its relationship to discourse comprehension. *Cognitive Science* 29:1045–60. [HK]
- Richardson, D. C., Dale, R. & Kirkham, N. Z. (2007) The art of conversation is coordination. *Psychological Science* 18:407–13. [HK, SOY]
- Richardson, M. J., Marsh, K. L., Isenhower, R. W., Goodman, J. R. L. & Schmidt, R. C. (2007) Rocking together: Dynamics of intentional and unintentional interpersonal coordination. *Human Movement Science* 26:867–91. [KJA, aMJJP]
- Ricœur, P. (1973) The model of the text: Meaningful action considered as a text. *New Literary History* 5:91–17. [GE]
- Riès, S., Janssen, N., Dufau, S., Alario, F.-X. & Burle, B. (2011) General-purpose monitoring during speech production. *Journal of Cognitive Neuroscience* 23:1419–36. [F-XA]
- Rizzolatti, G. & Sinigaglia, C. (2007) Mirror neurons and motor intentionality. *Functional Neurology* 22(4):205–10. [MP]
- Roche, J., Dale, R., Kreuz, R. J., & Jaeger, T. F. (2013) Learning to avoid syntactic ambiguity. Ms., University of Rochester. [TFJ]
- Roelofs, A. (2004) Error biases in spoken word planning and monitoring by aphasic and nonaphasic speakers: Comment on Rapp and Goldrick (2000). *Psychological Review* 111:561–72. [aMJJP]
- Rogalsky, C. & Hickok, G. (2011) The role of Broca’s area in sentence comprehension. *Journal of Cognitive Neuroscience* 23(7):1664–80. [ASD]
- Rommers, J., Meyer, A. S., Praamstra, P., & Huettig, F. (2013). The contents of predictions in sentence comprehension: Activation of the shape of objects before they are referred to. *Neuropsychologia* 51(3):437–47. [NM]
- Ross, E. D. (1981) The aprosodias: Functional-anatomical organization of the affective components of language in the right hemisphere. *Archives of Neurology* 38:561–68. [KJA]
- Rowland, C., Chang, F., Ambridge, B., Pine, J. M. & Lieven, E. V. (2012) The development of abstract syntax: Evidence from structural priming and the lexical boost. *Cognition* 125(1): 49–63. [FC]
- Rumelhart, D. E., Hinton, G. E. & Williams, R. J. (1986) Learning representations by back-propagating errors. *Nature* 323(6088):533–36. [FC]
- Rumiati, R. I., Zanini, S., Vorano, L. & Shallice, T. (2001) A form of ideational apraxia as a defective deficit of contentment scheduling. *Cognitive Neuropsychology* 18(7):617–42. DOI: 10.1080/02643290126375. [MP]
- Sacks, H., Schegloff, E. A. & Jefferson, G. (1974) A simplest systematics for the organization of turn-taking for conversation. *Language* 50:696–735. [aMJJP, HR]
- Saffran, J. R. (2002) Constraints on statistical language learning. *Journal of Memory and Language* 47:172–96. [MAJ]
- Saffran, J. R., Aslin, R. N. & Newport, E. L. (1996) Statistical learning by 8-month-old infants. *Science* 274:1926–28. [MAJ]
- Sahin, N. T., Pinker, S., Cash, S. S., Schomer, D. & Halgren, E. (2009) Sequential processing of lexical, grammatical, and articulatory information within Broca’s area. *Science* 326:445–49. [aMJJP]
- Salverda, A. P., Dahan, D. & McQueen, J. (2003) The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition* 90:51–89. [AMT]
- Sams, M., Möttönen, R. & Silvonen, T. (2005) Seeing and hearing others and oneself talk. *Brain Research: Cognitive Brain Research* 23(2–3):429–35. [KJA, GH, aMJJP]
- Sanford, A. J. & Garrod, S. C. (1981) *Understanding written language*. Wiley. [aMJJP]
- Sanford, A. J. & Sturt, P. (2002) Depth of processing in language comprehension: Not noticing the evidence. *Trends in Cognitive Sciences* 6:382–86. [SMM]
- Sartori, L., Beccio, C., Bara, B. G. & Castiello, U. (2009) Does the intention to communicate affect action kinematics? *Consciousness and Cognition* 18(3):766–72. DOI: 10.1016/j.concog.2009.06.004. [GP]
- Sato, M., Buccino, G., Gentilucci, M. & Cattaneo, L. (2010) On the tip of the tongue: Modulation of the primary motor cortex during audiovisual speech perception. *Speech Communication* 52(6):533–41. [ASD]
- Sato, M., Tremblay, P. & Gracco, V. L. (2009) A mediating role of the premotor cortex in phoneme segmentation. *Brain and Language* 111(1):1–7. DOI:10.1016/j.bandl.2009.03.002. [ASD]
- Schegloff, E. A. (1992) Repair after next turn: The last structurally provided defense of intersubjectivity in conversation. *American Journal of Sociology* 97(5): 1295–345. [CH]
- Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T. & Voegeley, K. (2013) Toward a second-person neuroscience. *Behavioral and Brain Sciences*. [KJA]
- Schippers, M. B., Roelbroeck, A., Renken, R., Nanetti, L. & Keysers, C. (2010) Mapping the information flow from one brain to another during gestural

- communication. *Proceedings of the National Academy of Science USA* 107:9388–93. [KJA]
- Schlenck, K.-J., Huber, W. & Willmes, K. (1987) “Prepairs” and repairs: Different monitoring functions in aphasic language production. *Brain and Language* 30:226–44. [aMJP]
- Schober, M. F. & Clark, H. H. (1989) Understanding by addressees and overhearers. *Cognitive Psychology* 21:211–32. [aMJP]
- Schriefers, H., Meyer, A. S. & Levelt, W. J. M. (1990) Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of Memory and Language* 29:86–102. [aMJP]
- Schwanenflugel, P. J. & Shoben, E. J. (1985) The influence of sentence constraint on the scope of facilitation for upcoming words. *Journal of Memory and Language* 24:232–52. [NM]
- Schwartz, J.-L., Basirat, A., Ménard, L. & Sato, M. (2012) The perception-for-action-control theory (PACT): A perceptuo-motor theory of speech perception. *Journal of Neurolinguistics* 25(5):336–54. [RL]
- Scott, S. & Johnsrude, I. S. (2003) The neuroanatomical and functional organisation of speech perception. *Trends in Neurosciences* 26: 100–107. [aMJP]
- Scott, S., McGettigan, C. & Eisner, F. (2009) A little more conversation, a little less action – candidate roles for the motor cortex in speech perception. *Nature Reviews Neuroscience* 10:295–302. [aMJP]
- Sebanz, N., Bekkering, H. & Knoblich, G. (2006a) Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences* 10(2):70–76. [GP, aMJP]
- Sebanz, N. & Knoblich, G. (2009) Prediction in joint action: What, when, and where. *Topics in Cognitive Science* 1:353–67. [aMJP]
- Sebanz, N., Knoblich, G., Prinz, W. & Wascher, E. (2006b) Twin peaks: An ERP study of action planning and control in coacting individuals. *Journal of Cognitive Neuroscience* 18:859–70. [aMJP]
- Segaert, K., Menenti, L., Weber, K., Petersson, K. M. & Hagoort, P. (2012) Shared syntax in language production and language comprehension – an fMRI study. *Cerebral Cortex* 22:1662–70. [aMJP]
- Shadmehr, R., Smith, M. A., & Krakauer, J. W. (2010) Error correction, sensory prediction, and adaptation in motor control. *Annual Review of Neuroscience* 33:89–108. [GH]
- Shapiro, L. (2011) *Embodied cognition*. Routledge. [GD]
- Sheehan, E. A., Namy, L. L. & Mills, D. L. (2007) Developmental changes in neural activity to familiar words and gestures. *Brain and Language* 101(3):246–59. [SK]
- Shintel, H. & Keysar, B. (2009) Less is more: A minimalist account of joint action in communication. *Topics in Cognitive Science* 1:260–73. [HK]
- Shockley, K., Santana, M. V. & Fowler, C. A. (2003) Mutual interpersonal postural constraints are involved in cooperative conversation. *Journal of Experimental Psychology: Human Perception and Performance* 29:326–32. [aMJP]
- Sidtis, J. J. & Sidtis, D. v. L. (2003) A neurobehavioural approach to dysprosody. *Seminars in Speech and Language* 24:93–105. [KJA]
- Skantze, G. & Hjalmarsson, A. (2010) Towards incremental speech generation in dialogue systems. *Proceedings of the 11th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, 1–8. [CH]
- Skipper, J. I., Goldin-Meadow, S., Nusbaum, H. C. & Small, S. L. (2007a) Speech-associated gestures, Broca’s area, and the human mirror system. *Brain and Language* 101(3):260–77. [ASD]
- Skipper, J. I., Goldin-Meadow, S., Nusbaum, H. C. & Small, S. L. (2009) Gestures orchestrate brain networks for language understanding. *Current Biology* 19:1–7. [ASD]
- Skipper, J. I., Nusbaum, H. C. & Small, S. L. (2005) Listening to talking faces: Motor cortical activation during speech perception. *NeuroImage* 25(1):76–89. [ASD]
- Skipper, J. I., van Wassenhove, V., Nusbaum, H. C. & Small, S. L. (2007b) Hearing lips and seeing voices: How cortical areas supporting speech production mediate audiovisual speech perception. *Cerebral Cortex* 17:2387–99. [ASD]
- Slevc, L. R. (2011) Saying what’s on your mind: Working Memory effects on sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 37(6):1503–14. [LRS]
- Smith, A. & Zelaznik, H. N. (2004) Development of functional synergies for speech motor coordination in childhood and adolescence. *Developmental Psychobiology* 45(1):22–33. DOI: 10.1002/dev.20009. [SK]
- Sommer, M. A. & Wurtz, R. H. (2008) Brain circuits for the internal monitoring of movements. *Annual Review of Neuroscience* 31:317–38. [F-XA]
- Sonderegger, M. & Yu, A. (2010) *A rational account of perceptual compensation for coarticulation*. Paper presented at the Proceedings of the 32nd Annual Meeting of the Cognitive Science Society (CogSci10). [TFJ]
- Sorace, A. (2011) Pinning down the concept of “interface” in bilingualism. *Linguistic Approaches to Bilingualism* 1:1–33. [JF]
- Stack, D. M. (2007) The salience of touch and physical contact during infancy: Unravelling some of the mysteries of the somesthetic sense. In: *Blackwell handbook of infant development*, ed. G. Bremner & A. Fogel, ch. 13. Blackwell. [KJA]
- Stanley, J., Gowen, E. & Miall, R. C. (2007) Effects of agency on movement interference during observation of a moving dot stimulus. *Journal of Experimental Psychology: Human Perception and Performance* 33:915–26. [aMJP]
- Staub, A. & Clifton, C., Jr. (2006) Syntactic prediction in language comprehension: Evidence from either...or. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 32:425–36. [aMJP]
- Staudte, M. & Crocker, M. W. (2011) Investigating joint attention mechanisms through spoken human-robot interaction. *Cognition* 120:268–91. [HK]
- Stephens, G. J., Silbert, L. J. & Hasson, U. (2010) Speaker-listener neural coupling underlies successful communication. *Proceedings of the National Academy of Sciences* 107:14425–30. [aMJP]
- Stern, D. N. (2010) *Forms of vitality: Exploring dynamic experience in psychology and the arts*. Oxford University Press. [KJA]
- Stevens, K. N. & Halle, M. (1967) Remarks on the analysis by synthesis and distinctive features. In: *Models for the perception of speech and visual form*, ed. W. Walthe-Dunn, pp. 88–102. MIT Press. [GH]
- Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., De Ruiter, J. P., Yoon, K. E., & Levinson, S. C. (2009) Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences of the United States of America* 106(26):10587–92. [JPdR]
- Straube, B., Green, A., Bromberger, B. & Kircher, T. (2011) The differentiation of iconic and metaphoric gestures: Common and unique integration processes. *Human Brain Mapping* 32(4):520–33. DOI: 10.1002/hbm.21041. [ASD]
- Strijkers, K. & Costa, A. (2011) Ridding the lexical speedway: A critical review on the time course of lexical selection in speech production. *Frontiers in Psychology* 2:356. [KS]
- Strijkers, K., Holcomb, P. & Costa, A. (2011) Conscious intention to speak facilitates lexical access during overt object naming. *Journal of Memory and Language* 65:345–62. [KS]
- Suttle, L. & Goldberg, A. E. (forthcoming) *Learning what not to say: Comparing the role of preemption and entrenchment*. Princeton University. [MAJ]
- Swinney, D. (1979) Lexical access during sentence comprehension: (Re) consideration of context effects. *Journal of Verbal Learning and Verbal Behavior* 18:645–59. [aMJP]
- Tanenhaus, M. K. (2007) Eye movements and spoken language processing. In: *Eye movements: A window on mind and brain*, ed. R. P. G. van Gompel, M. H. Fischer, W. S. Murray & R. L. Hill, pp. 309–26. Elsevier. [LRS]
- Tettamanti, M., Buccino, G., Saccuman, M. C., Gallese, V., Danna, M., Scifo, P., Fazio, F., Rizzolatti, G., Cappa, S. F. & Perani, D. (2005) Listening to action-related sentences activates fronto-parietal motor circuits. *Journal of Cognitive Neuroscience* 17(2):273–81. DOI: 10.1162/0899929053124965. [MP]
- Thompson-Schill, S. L., Jonides, J., Marshuetz, C., Smith, E. E., D’Esposito, M., Kan, I. P., Knight, R. T. & Swick, D. (2002) Effects of frontal lobe damage on interference effects in working memory. *Cognitive, Affective & Behavioral Neuroscience* 2(2):109–20. [LRS]
- Tian, X. & Poeppel, D. (2010) Mental imagery of speech and movement implicates the dynamics of internal forward models. *Frontiers in Psychology* 1:166. [MIM, aMJP, KS]
- Tomasello, M. (2008) *Origins of human communication*. Bradford Books/MIT Press. [KJA]
- Tomasello, M., Carpenter, M., Call, J., Behne, T. & Moll, H. (2005) Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences* 28:675–91. [HK]
- Toni, I., de Lange, F. P., Noordzij, M. L. & Hagoort, P. (2008) Language beyond action. *Journal of Physiology – Paris* 102(1–3):71–79. DOI: 10.1016/j.jphysparis.2008.03.005. [GD, MP]
- Tourville, J. A. & Guenther, F. H. (2011) The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes* 26:952–81. [F-XA, aMJP]
- Tourville, J. A., Reilly, K. J. & Guenther, F. K. (2008) Neural mechanisms underlying auditory feedback control of speech. *NeuroImage* 39:1429–43. [JPdR, MIM, aMJP, KS]
- Towle, V. L., Yoon, H. A., Castelle, M., Edgar, J. C., Biassou, N. M., Frim, D. M., Spire, J. P. & Kohrman, M. H. (2008) ECoG gamma activity during a language task: Differentiating expressive and receptive speech areas. *Brain* 131:2013–27. [F-XA]
- Tremblay, P., Sato, M. & Small, S. L. (2012) TMS-induced modulation of action sentence priming in the ventral premotor cortex. *Neuropsychologia* 50(2):319–26. DOI:10.1016/j.neuropsychologia.2011.12.00. [ASD]
- Tremblay, P. & Small, S. L. (2011) On the context-dependent nature of the contribution of the ventral premotor cortex to speech perception. *NeuroImage* 57(4):1561–71. [ASD]
- Tremblay, S., Shiller, D. M. & Ostry, D. J. (2003) Somatosensory basis of speech production. *Nature* 423: 866–69. [rMJP]
- Tretriluxana, J., Gordon, J. & Winstein, C. J. (2008) Manual asymmetries in grasp pre-shaping and transport-grasp coordination. *Experimental Brain Research* 188(2):305–15. DOI: 10.1007/s00221-008-1364-2. [MP]
- Trevarthen, C. & Aitken, K. J. (2001) Infant intersubjectivity: Research, theory, and clinical applications. *Annual Research Review. Journal of Child Psychology and Psychiatry* 42:3–48. [KJA]

- Trude, A. M. & Brown-Schmidt, S. (2012) Talker-specific perceptual adaptation during on-line speech perception. *Language and Cognitive Processes* 27: 979–1001. [rMJP, AMT]
- Trueswell, J. C. & Tanenhaus, M. K. (ed.) (2005) *Processing world-situated language: Bridging the language-as-action and language-as-product traditions*. MIT Press. [SOY]
- Trueswell, J. C., Tanenhaus, M. K. & Garnsey, S. M. (1994) Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language* 33:285–318. [aMJJP]
- Tuomela, R. (2007) *The philosophy of sociality*. Oxford University Press. [YK]
- Turk-Browne, N. B., Scholl, B. J., Johnson, M. K. & Chun, M. M. (2010) Implicit perceptual anticipation triggered by statistical learning. *Journal of Neuroscience* 30:11177–87. [MAJ]
- Tversky, A. & Kahneman, D. (1973) Availability: A heuristic for judging frequency and probability. *Cognitive Psychology* 5(2):677–95. [NM]
- Umiltà, M. A., Kohler, E., Gallese, V., Fogassi, L., Fadiga, L., Keysers, C. & Rizzolatti, G. (2001) I know what you are doing: A neurophysiological study. *Neuron*, 32:91–101. [aMJJP]
- Ungerleider, L. G. & Haxby, J. V. (1994) “What” and “where” in the human brain. *Current Opinion in Neurobiology* 4(2):157–65. [ASD]
- Van Berkum, J. J. A., Brown, M. C., Zwitserlood, P., Kooijman, V. & Hagoort, P. (2005) Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 31:443–67. [NM, aMJJP]
- Van Berkum, J. J. A., van den Brink, D., Tesink, C. M. J. Y., Kos, M. & Hagoort, P. (2008) The neural integration of speaker and message. *Journal of Cognitive Neuroscience* 20:580–91. [HK]
- Van den Bussche, E., Van den Noortgate, W. & Reynvoet, B. (2009) Mechanisms of masked priming: A meta-analysis. *Psychological Bulletin* 135:452–77. [aMJJP]
- Van Lancker, D. & Breitenstein, C. (2000) Emotional dysprosody and similar dysfunctions. In: *Behavior and mood disorders in focal brain lesions*, ed. J. Bogousslavsky & J. L. Cummings, ch. 12. Cambridge University Press. [KJA]
- Van Overvalle, F. & Baetens, K. (2009) Understanding others’ actions and goals by mirror and mentalizing systems: A meta-analysis. *NeuroImage* 48(3):564–84. DOI: 10.1016/j.neuroimage.2009.06.009. [MP]
- Van Schie, H. T., van Waterschoot, B. M. & Bekkering, H. (2008) Understanding action beyond imitation: Reversed compatibility effects of action observation in imitation and joint action. *Journal of Experimental Psychology: Human Perception and Performance* 34:1493–500. [aMJJP]
- van Wassenhove, V., Grant, K. W. & Poeppel, D. (2005) Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences* 102(4):1181–86. [ASD, GH]
- Van Wijk, C. & Kempen, G. (1987) A dual system for producing self-repairs in spontaneous speech: Evidence from experimentally elicited corrections. *Cognitive Psychology* 19:403–40. [aMJJP]
- Venezia, J. H., Saberi, K., Chubb, C. & Hickok, G. (2012) Response bias modulates the speech motor system during syllable discrimination. *Frontiers in Psychology* 3. article 157, doi: 10.3389/fpsyg.2012.00157 [GH]
- Vesper, C., van der Wel, R. P. R. D., Knoblich, G. & Sebanz, N. (2011) Making oneself predictable: Reduced temporal variability facilitates joint action coordination. *Experimental Brain Research* 211(3–4):517–30. DOI:10.1007/s00221-011-2706-z. [GP]
- Vigliocco, G., Antonini, T. & Garrett, M. F. (1997) Grammatical gender is on the tip of Italian tongues. *Psychological Science* 8:314–17. [aMJJP]
- Vigliocco, G. & Hartsuiker, R. J. (2002) The interplay of meaning, sound, and syntax in sentence production. *Psychological Bulletin* 128:442–72. [aMJJP]
- Vigneau, M., Beaucousin, V., Hervé, P. Y., Duffau, H., Crivello, F., Houdé, O., Mazoyer, B. & Tzourio-Mazoyer, N. (2006) Meta-analyzing left hemisphere language areas: Phonology, semantics, and sentence processing. *NeuroImage* 30(4):1414–32. DOI: 10.1016/j.neuroimage.2005.11.002. [KJA, ASD, aMJJP]
- Visser, C. T., Chwilla, D. J. & Kolk, H. H. (2006) Monitoring in language perception: The effect of misspellings of words in highly constrained sentences. *Brain Research* 1106:150–63. [aMJJP]
- Vygotsky, L. S. (1962) *Thought and language* (E. Hanfmann & G. Vakar, trans.). MIT Press. [GMO]
- Warker, J. A. & Dell, G. S. (2006) Speech errors reflect newly learned phonotactic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 32(2):387. [TFJ]
- Watkins, K. & Paus, T. (2004) Modulation of motor excitability during speech perception: The role of Broca’s area. *Journal of Cognitive Neuroscience* 16:978–87. [aMJJP]
- Watkins, K., Strafella, A. P. & Paus, T. (2003) Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia* 41:989–94. [ASD, aMJJP]
- Weber, A., Grice, M. & Crocker, M. W. (2006) The role of prosody in the interpretation of structural ambiguities: A study of anticipatory eye movements. *Cognition* 99:B63–72. [aMJJP]
- Wei, K. & Kording, K. (2009) Relevance of error: What drives motor adaptation? *Journal of Neurophysiology* 101(2):655–64. [TFJ]
- Wheeldon, L. R. & Levelt, W. J. M. (1995) Monitoring the time course of phonological encoding. *Journal of Memory and Language* 34:311–34. [aMJJP]
- Wijnen, F. & Kolk, H. H. J. (2005) Phonological encoding, monitoring, and language pathology: Conclusions and prospects. In: *Phonological encoding in normal and pathological speech*, ed. R. J. Hartsuiker, R. Bastiaanse, A. Postma & F. Wijnen, pp. 283–304. Psychology Press. [aMJJP]
- Wilkes-Gibbs, D. & Clark, H. H. (1992) Coordinating beliefs in conversation. *Journal of Memory and Language*, 31:183–94. [SOY]
- Willems, R. M., Özyürek, A. & Hagoort, P. (2007) When language meets action: The neural integration of gesture and speech. *Cerebral Cortex* 17(10):2322. [ASD]
- Willems, R. M., Özyürek, A. & Hagoort, P. (2009) Differential roles for left inferior frontal and superior temporal cortex in multimodal integration of action and language. *NeuroImage* 47:1992–2004. [ASD]
- Wilson, M. (2002) Six views of embodied cognition. *Psychonomic Bulletin and Review* 9:625–36. [GD]
- Wilson, M. & Knoblich, G. (2005) The case for motor involvement in perceiving conspecifics. *Psychological Bulletin* 131:460–73. [aMJJP]
- Wilson, M. & Wilson, T. P. (2005) An oscillator model of the timing of turn-taking. *Psychonomic Bulletin & Review* 12:957–68. [aMJJP]
- Wilson, S. M. & Iacoboni, M. (2006) Neural responses to non-native phonemes varying in producibility: Evidence for the sensorimotor nature of speech perception. *Neuroimage* 33(1):316–25. [ASD, GH]
- Wilson, S. M., Saygin, A. P., Sereno, M. I. & Iacoboni, M. (2004) Listening to speech activates motor areas involved in speech production. *Nature Neuroscience* 7(7):701–702. [ASD, aMJJP]
- Wohlschläger, A. (2000) Visual motion priming by invisible actions. *Vision Research* 40:925–30. [aMJJP]
- Wolpert, D. M. (1997) Computational approaches to motor control. *Trends in Cognitive Sciences* 1:209–16. [MAJ, MIM, aMJJP]
- Wolpert, D. M., Diedrichsen, J. & Flanagan, J. R. (2011) Principles of sensorimotor learning. *Nature Reviews Neuroscience* 12:739–51. [FC]
- Wolpert, D. M., Doya, K. & Kawato, M. (2003) A unifying computational framework for motor control and social interaction. *Philosophical Transactions of the Royal Society B: Biological Sciences* 358(1431):593–602. DOI:10.1098/rstb.2002.1238. [HK, GP, aMJJP]
- Wolpert, D. M., Ghahramani, Z. & Flanagan, J. R. (2001) Perspectives and problems in motor learning. *Trends in Cognitive Sciences* 5:487–94. [MAJ, arMJJP]
- Wolpert, D. M., Ghahramani, Z. & Jordan, M. I. (1995) An internal model for sensorimotor integration. *Science* 269(5232):1880–82. [GH]
- Wrangham, R. (2009) *Catching fire: How cooking made us human*. Basic Books. [KJA]
- Wright, B. & Garrett, M. F. (1984) Lexical decision in sentences: Effects of syntactic structure. *Memory & Cognition* 12:31–45. [aMJJP]
- Xu, J., Gannon, P. J., Emmorey, K., Smith, J. F. & Braun, A. R. (2009) Symbolic gestures and spoken language are processed by a common neural system. *Proceedings of the National Academy of Sciences* 106(49):20664–69. [ASD]
- Xu, S., Zhang, S. & Geng, H. (2011) Gaze-induced joint attention persists under high perceptual load and does not depend on awareness. *Vision Research* 51:2048–56. [HK]
- Yoon, S. O., Koh, S. & Brown-Schmidt, S. (2012) Influence of perspective and goals on reference production in conversation. *Psychonomic Bulletin & Review* 19:699–707. [SOY]
- Yoshida, M., Dickey, M. W. & Sturt, P. (2013) Predictive processing of syntactic structure: Sluicing and ellipsis in real-time sentence processing. *Language and Cognitive Processes* 28:272–302. [aMJJP]
- Yuen, I., Davis, M. H., Brysbaert, M. & Rastle, K. (2010) Activation of articulatory information in speech perception. *Proceedings of the National Academy of Sciences* 107:592–97. [aMJJP]
- Zekveld, A. A., Heslenfeld, D. J., Festen, J. M. & Schoonhoven, R. (2006) Top-down and bottom-up processes in speech comprehension. *NeuroImage* 32:1826–36. [RL]
- Zlatev, J. (2008) From proto-mimesis to language: Evidence from primatology and social neuroscience. *Journal of Physiology-Paris* 102(1–3):137–51. DOI: 10.1016/j.jphysparis.2008.03.016. [MP]