



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Predicting and imagining language

Citation for published version:

Gambi, C & Pickering, M 2015, 'Predicting and imagining language' *Language, Cognition and Neuroscience*, pp. 1-13. DOI: 10.1080/23273798.2015.1049188

Digital Object Identifier (DOI):

[10.1080/23273798.2015.1049188](https://doi.org/10.1080/23273798.2015.1049188)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Language, Cognition and Neuroscience

Publisher Rights Statement:

This is an Accepted Manuscript of an article published by Taylor & Francis Group in *Language, Cognition and Neuroscience* on 03/09/2015, available online:
<http://www.tandfonline.com/doi/full/10.1080/23273798.2015.1049188#.Ve2vUZe2pqw>.

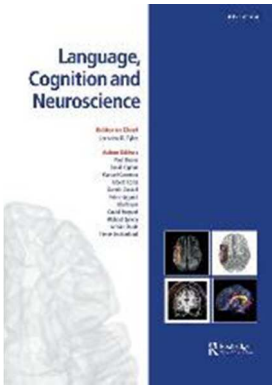
General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.





Predicting and imagining language

Journal:	Language, Cognition and Neuroscience
Manuscript ID:	PLCP-2014-PredLCP-9744.R1
Manuscript Type:	Special Issue
Date Submitted by the Author:	n/a
Complete List of Authors:	Gambi, Chiara; University of Edinburgh, Pickering, Martin; University of Edinburgh,
Keywords:	speech imagery, prediction, forward models, joint tasks

SCHOLARONE™
Manuscripts

RUNNING HEAD: Prediction and imagination

Predicting and imagining language

Chiara Gambi^{a*} and Martin J. Pickering^a

^a Department of Psychology, University of Edinburgh, Edinburgh, U.K.

WORD COUNT: 9308 (including references)

To what extent is predicting language akin to imagining language? Recently, researchers have argued that covert simulation of the production system underlies both articulation imagery and predicting what somebody is about to say. Moreover, experimental evidence implicates potentially similar production-related mechanisms in prediction during language comprehension and in mental imagery tasks. We discuss evidence in favour of this proposal, and argue that imagining others' utterances can also implicate covert simulation. Finally, we briefly review evidence that speakers in joint language tasks cannot help but mentally represent (i.e., imagine) whether others are engaging in language production, and that they do so using mechanisms that are also implicated in preparing to speak.

Keywords: speech imagery, prediction, forward models, joint tasks

*Correspondence:

Chiara Gambi

Department of Psychology

University of Edinburgh

Edinburgh

United Kingdom

cgambi@exseed.ed.ac.uk

Contact details for other authors:

Martin.pickering@ed.ac.uk

RUNNING HEAD: Prediction and imagination

Introduction

To what extent is predicting language akin to imagining language? When I am in my favourite café ordering dessert, while saying *I will have a* I can **predict** that I am about to say *Achocolate and orange panna cotta, please*. But if I am abroad and therefore unable to go to my favourite café, I can also **imagine** myself saying *Achocolate and orange panna cotta, please* just as well as I can imagine eating one. So I can both predict and imagine my own utterances.

I can also predict and imagine somebody else's utterances. When listening to a child, for example, I can predict that the utterance *I want to have a* is more likely to be followed by *sweet* than by *swede*. Importantly, I can do this very quickly, over the course of a few hundred milliseconds, which means that I can predict the child's utterance in real-time. This allows me, for example, to look at a jar full of sweets even before the child utters the word *sweet* (e.g., Altmann & Kamide, 1999). In addition if I know a child is in the next-door room, and she has just called her mum in to ask for something, I am more likely to imagine the child saying *I want a sweet* than *I want a swede*.

For the purposes of this paper, we regard imagination (or imagery) of utterances as any mental representation of an utterance that is not involved in an instance of production or an instance of comprehension.¹ The content of such mental representation can be both perceptual (e.g., what articulating *panna cotta* or *sweet* feels like and what hearing *panna cotta* or *sweet* sounds like) and conceptual (e.g., what using a polite utterance form is like). This definition of imagination is ~~very much related~~ akin to Goldman's notion of enactment-imagination (or E-imagination; Goldman, 2006; Chapter 7), which corresponds to our ability to recreate our own or others' mental states (e.g., seeing, believing, or desiring) in our mind.

In this paper we propose that imagining and producing language are directly related. In the first section, we argue that imagination is a form of offline prediction. But predicting oneself and predicting other people are distinct, and imagining oneself and imagining other people are likely to be distinct as well. We therefore contrast how people predict themselves

RUNNING HEAD: Prediction and imagination

versus other people, and then consider different forms of other-prediction. In each case, we argue that there is a parallel way to imagine language. The next two sections consider in more detail the two principal ways in which we imagine other people's language: by attributing them an intention and then using our own production mechanisms (imagination-by-simulation); and by treating their language like any regular but non-intentional object (imagination-by-association). The final section applies the account to the imagination of dialogue and the imagination of other people's utterances in joint tasks.

Imagination as offline prediction

Given our definition of imagination, there are several similarities between prediction and imagination. First, in both cases, I rely on my previous experience with my own or with children's language and on my knowledge of socio-cultural conventions about ordering food in restaurants and about children's behaviour. Thus, the kind of information on which I base predictions and imagination is (potentially) the same. Second, in both cases, I internally generate a representation of what I am about to say (*chocolate and orange panna cotta*) or the child is about to say (*sweet*) without having heard myself or the child saying it. Third, in both cases, this internal representation could either match or mismatch what I or the child will actually end up saying. For example, I might get distracted and say *chocolate and strawberry panna cotta* (because my eyes happen to be looking at that item on the menu) by mistake; or the child might ask for a toy and not a sweet. Alternatively, I or the child might end up not speaking at all.

Of course, there are also differences between the two situations. Most obviously, I have a wealth of information available in the immediate context during prediction that is not available in imagination. For instance, I can see and hear the child and the context she is in (which might include sweets and not swedes). This contextual information is likely to affect my accuracy – my representation is more likely to match “the outside world” during prediction than imagination. But this accuracy disadvantage for imagination does not

RUNNING HEAD: Prediction and imagination

necessarily always hold true – I could predict badly if I misjudge the child, or I could imagine very accurately, if I have accurate knowledge of the child and the situation (the contents of the next-door room). However, these differences do not constitute principled reasons why the cognitive mechanisms underlying prediction and imagination need to be distinct.

A more interesting difference is that I can compare my internal representation to reality in prediction but I cannot (typically) do this in imagination. And I am presumably aware of this difference while I am predicting or imagining. To illustrate, if I think about ordering a panna cotta on my plane home, am I predicting what I will do or imagining what I would be doing if I were at home? What distinguishes prediction from imagination in this case seems to be the type of propositional attitude I direct towards the external world. When I predict what I will do, I hold a belief about the state of the world in the future, and I am aware that I will be able to compare my belief to my perception of that future state and determine whether my belief matches the world. When I imagine what I would be doing if I were home, instead, it is irrelevant whether my internal state matches the state of the world (either now or at a later point in time). But what if I am uncertain whether I will get to the café before it closes? Does it mean that I oscillate between predicting and imagining? In fact, we argue that the eventual state of the world (whether the café remains open or not) is not relevant for establishing what types of processes are involved: Even if I construct a different propositional attitude when I believe I am predicting versus imagining, the processes involved appear closely related.

In the thinking-on-a-plane example above, the distinction between prediction and imagination is difficult because the focus is on a non-prototypical instance of prediction. The instance is non-prototypical because I am predicting what will happen hours into the future. Instead, prototypical cases of prediction, we maintain, are those in which I predict on a much more rapid time-scale, or even “in real time”. Crucially, much prediction of language (as well as other things, such as visual stimuli) is indeed very rapid, as revealed by several eye-tracking and electrophysiological studies of language comprehension (e.g., Altmann & Kamide, 1999; Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005; see Kutas,

RUNNING HEAD: Prediction and imagination

De Long, & Smith, 2011 for review). For example, in a ~~elassie~~ visual-world eye-tracking study, Altmann and Kamide (1999) had listeners hear the sentence *The boy will eat the cake* while looking at a scene containing only one edible object (a picture of a cake). Within 600 ms from the onset of *eat* (and 85 ms before the onset of *cake*) participants began launching predictive eye-movements towards the cake. Similarly, using electrophysiological measures Van Berkum et al. (2005; Exp. 1) showed that Dutch listeners are sensitive to the mismatch between the gender of an highly expected noun and the gender-marked inflection of a prenominal adjective that precedes the noun by around 700 ms.

Predictions might be computed at even shorter time-scales during the production of individual sounds. For instance, if auditory feedback is unaltered, the N100 (or M100 in MEG studies) is typically suppressed when a participant is speaking compared to when the participant is listening to a recording of his/her voice. But when feedback is modified (e.g., by shifting pitch), such suppression does not occur, suggesting that speakers are sensitive to the mismatch between intended and heard speech sounds within 100 ms ~~from-of~~ sound onset (e.g., Heinks-Maldonado, Nagarajan, & Houde, 2006; Behroozmand, & Larson, 2011).

Instances of online prediction in language comprehension and language production appear to be quite distinct from imagination (and perhaps from long-term prediction such as what kind of language I will hear on the radio tomorrow morning, or what jokes I will make during my lecture in a week's time). The point is that in instances of online prediction, predictive representations of my own or others' utterances are used to support either the overt production or the comprehension of language. Instead, when I imagine my own or others' utterances, these internal representations are decoupled from overt production and comprehension. Importantly, this does not mean that they cannot influence overt production or comprehension processes (or be influenced by such processes). However, representations formed during imagination do not serve a functional role in production or comprehension.

Interestingly, Goldman (2006) also draws a distinction between online processes and imagination. However, he goes further and suggests that the mechanisms involved in online processing are entirely unrelated to the mechanisms involved in imagination. According to

RUNNING HEAD: Prediction and imagination

Goldman, for example, when I imagine moving my hand, a certain mental state (i.e., the sensation of hand movement) is recreated in the absence of actual movement (i.e., offline). But while I actually move my hand (i.e., online) no such process takes place. In contrast, others have argued that what drives imagination is precisely the same process underlying prediction during actual movement (Grush, 2004; Hurley, 2008). We agree with the latter position. Crucially, once one recognizes that online prediction also involves the “creation” of mental states (e.g., the predicted sensation of hand movement), the hypothesis that online prediction and offline imagination are related becomes likely.

Our proposal is that people primarily use predictive mechanisms to predict language “online”. These mechanisms can then be “extracted” to make off-line predictions and specifically for imagination, when people are not trying to predict what might be going to happen. But we propose that the mechanisms are not fundamentally altered by being extracted from online processing in this way. In a related proposal, Pezzulo (2011) argued that the ability to take prediction off-line gave rise, over the course of human evolution, to internal representations that can be used in the absence of overt action and perception. Our suggestion is superficially similar, but we make no claim regarding the development of linguistic representations. Rather, we suggest that the set of mechanisms used for predicting our own and others’ utterances in real time can be also employed to imagine those utterances while they are not being articulated or comprehended. More specifically, we now propose that people make use of different forms of prediction, and argue that these are mirrored in different forms of imagination. The first distinction is between self- and other-prediction; the second is between other-prediction based on production and other-prediction based on comprehension. [The language prediction mechanisms discussed in this paper are largely based on Pickering and Garrod \(2013\), and our main purpose is to discuss how their proposal can be applied to imagination. Since Pickering and Garrod’s proposal is quite recent, we describe it in some detail below.](#)

Predicting ~~and imagining~~ one’s own utterances

RUNNING HEAD: Prediction and imagination

According to Pickering and Garrod (2013), when I prepare to speak I predict myself in the following way (Figure 1A, *Self-prediction*). I formulate an intention to communicate, which roughly consists of a situation model and a representation of the speech act (statement, question, etc.), plus a “command” to initiate the utterance. This intention is pre-linguistic, meaning that it does not specify the linguistic properties of the utterance that will be produced (e.g., whether it will be active or passive, or what sounds it will contain). However, I then can use learned intention-to-production mappings to predict some aspects of what it would be like to produce an utterance expressing that intention. For example, if I have regularly had a similar intention before (e.g., to order a dessert in a café), I might predict that I will utter a polite question (e.g., “may I have”) – as I have (almost) always done so whenever I have had that intention in the past. Or if I have a more detailed intention (e.g., I want to order something chocolaty but fruity), I can predict a meaning representation corresponding to the concept FRUIT, or even the initial sound /t/. Again, these predictions are based on my behaviour during previous occurrences of similar intentions (e.g., when I want chocolate. I tend to produce the sound /t/).

These mappings are computed using what Pickering and Garrod (2013) term a forward production model. This is a model of the process of language production: It captures my (implicit) knowledge about the way such process translates intentions into movements of the articulators. The output of the forward production model is an (approximate) representation of the state of the production system at various processing levels (e.g., phonology: what phoneme will become activated first?). Then, I use a forward comprehension model to compute mappings between production and comprehension representations, and predict some aspects of what it would be like to comprehend myself producing an utterance expressing that intention (e.g., I will experience the sound /t/). The forward comprehension model is a model of the process of language comprehension, and it captures (implicit) knowledge about how this process translates sounds into communicative intentions. These forward models constitute components of the production system, but they

RUNNING HEAD: Prediction and imagination

do not by themselves lead to overt production. In the terminology of Pickering and Clark (2014), they are auxiliary forward models.

Of course, the speaker's intention is primarily used to generate an implemented series of production and comprehension representations associated with preparing speech (Figure 1A, *Self-prediction*, top arrows). These representations are constructed over time, for example with representations of meaning (e.g., the concepts CHOCOLATE, ORANGE) unfolding before representations of sound (the word forms /'ʃɒkəlɪt/, /'ɒrɪndʒ/). The predicted production and comprehension representations (Figure 1A, *Self-prediction*, bottom arrows), which are the output of the forward models, are ready well before the implemented representations (for example e.g., the predicted sound representation /tʃ/ could be ready well before the implemented sound representation /'ʃɒkəlɪt/). Moreover predicted representations are constructed at roughly the same time, and can then be compared with the implemented representations as soon as those representations become available. For example, I compare the implemented meaning representation and the predicted meaning representation (in say 300ms) and then compare the implemented sound representation and the predicted sound representation (in say 600ms). If I start to construct the meaning STRAWBERRY, I notice the anomaly earlier (just after 300ms); if I construct the right meaning but make a sound error (e.g., select the sound /ʃ/ instead of /tʃ/ because the waiter, who has a very strong French accent, has just said /'ʃɒkəlɪt/), I notice the anomaly later (just after 600ms).

According to Pickering and Garrod (2013), speakers compare actual and predicted comprehension-based representations (monitor box to the right in Figure 1A, *Self-prediction*), and comprehension-based representations always follow production-based representations (Pickering and Garrod do not consider comparison between predicted and actual production-based representations; but see Nozari, Dell, & Schwartz, 2011). However, a slightly different account allows for construction of predicted comprehension-based representations from the production command, via a direct mapping that does not require building predicted production-based representations, so that predicted comprehension and production-based representations could be built at the same time. In fact, if production and comprehension

RUNNING HEAD: Prediction and imagination

share representations at the semantic and syntactic levels (Kempen, 2014; Pickering & Garrod, 2004), production-to-comprehension mappings [at these levels](#) might be trivial to compute in practice ([Gambi & Pickering, forthcoming](#)).

Much of the evidence for self-prediction comes from self-monitoring. For example, Pickering and Garrod (2013, 2014) pointed out that traditional comprehension-based monitoring (Levelt, 1989) is likely to be too slow to account for self-correction and therefore suggested that people make their predictions (at different linguistic levels) and compare those predictions (specifically, the predicted comprehension-based representations) against the utterance as they prepare it. This means that they can detect (and potentially begin to correct) errors as soon as they formulate (or “implement”) a level of representation (e.g., of phonology). As mentioned in the previous section, the earliest auditory component in the EEG and MEG records is reduced when speakers hear their own unaltered voice but not when their voice is modified ([e.g., Heinks-Maldonado et al., 2006](#)). Crucially, even with unaltered feedback, the degree of suppression is a function of the deviation of the self-produced sound from the speaker’s own “average” production (Niziolek, Nagarajan, & Houde, 2013). This suggests that speakers predict what they should sound like, as assumed by computational models of speech motor control (e.g., Hickok, 2012a). Importantly, in these models, self-predictions are computed using forward models. Evidence for self-prediction at other linguistic levels is rather more indirect (see Pickering & Garrod, 2013; p. 340).

INSERT FIGURE 1 ABOUT HERE

Imagining one’s own utterances

~~Our~~ [The first part of our](#) proposal is that forward production and forward comprehension models, as used during self-prediction, might be responsible for me imagining myself saying *panna cotta*. This constitutes an extension of Pickering and Garrod’s (2013) model of prediction during production of one’s own utterances to the imagination of one’s own utterances. We propose that people can imagine utterances simply by “extracting” the

RUNNING HEAD: Prediction and imagination

forward production and forward comprehension models and making use of their output to determine what they would say. See Figure 1A, for the process of extraction (shaded area). Most likely, they would imagine either the meaning or the phonological (or phonetic) form of the utterance (or both). They could then use the predicted representations to keep track of what they have imagined, and also to construct a further intention and continue imagining. Such self-imagination is therefore dependent on the forward models, which constitutes a component of the production system.

This proposal about self-imagination is related to Tian and Poeppel's (2010, 2012) suggestion that some of the same neural processes take place when I prepare to articulate a syllable as when I imagine myself articulating that syllable. According to these authors, when I imagine myself producing the syllable [da], I use a forward model to generate predictions about what producing that syllable would feel like (imagined somatosensory feedback) and sound like (imagined auditory feedback).² This causes activation in areas of the parietal cortex that are also activated when preparing to produce the same syllable and in areas of the auditory cortex that are also activated when listening to the same syllable (Tian & Poeppel, 2010), thereby suggesting that articulation imagery involves a mapping from motor to somatosensory and from somatosensory to auditory representations. Crucially, the activation in auditory cortex occurs very rapidly and around the time at which it occurs during overt speech (~ 170ms), which suggests that it might be caused by the same set of mappings that underlies self-prediction during preparation for speech.

Note that two prominent models of speech motor control assume that motor-to-somatosensory and motor-to-auditory mappings occur in parallel rather than sequentially (Hickok, 2012a; Tourville & Guenther, 2011). While proposing a sequential architecture, Tian and Poeppel (2012) also admitted the possibility of direct motor-to-auditory mappings, especially when the system is under pressure to compute mappings quickly (p. 5). But even if this parallel alternative is correct, a recent MEG study (Tian & Poeppel, 2013) showed that overt articulation and articulation imagery of a syllable had equivalent effects on activation of primary auditory cortex when the same syllable was subsequently presented to participants

RUNNING HEAD: Prediction and imagination

via headphones (compared to when a different syllable was presented). This indicates that overt articulation and covert (imagined) articulation both enhance subsequent neural responses to the same sound. Moreover, it suggests that shared neural mechanisms could be responsible for the computation of predicted sensory consequences during articulation imagery and during speech, although it does not compellingly demonstrate the latter.

Further suggestive evidence in this direction comes from sign language. For people using sign language, visual information is available for signs produced by other people but only to a limited extent for self-produced signs. This is because most signs are produced at the periphery of the visual field. As a consequence, signing is not negatively affected when visual information is degraded or removed (Emmorey, Bosworth, & Kraljic, 2009, Exp. 2), in contrast with speech production, which is strongly affected by distorted auditory feedback. This suggests that signers rely mostly on somatosensory feedback during self-monitoring. Interestingly, deaf patients with schizophrenia seem to experience somatosensory images during hallucinations of “signing” more than hearing schizophrenic patients do during hallucinations of “voices” (Atkinson, 2006), which is consistent with the idea that mechanisms involved in self-monitoring are also implicated in the generation of such images.

More compellingly, Tian and Poeppel (2015) recently showed that M100 is suppressed when a heard vowel matches a covertly (i.e., without any movement of the articulators) produced vowel. This result mirrors M100 suppression that is known to occur when people speak overtly. Therefore, it strongly suggests that overt articulation and covert articulation (what they term articulation imagery) make use of very similar mechanisms. Further evidence that such shared mechanism is a simulation mechanism comes from the fact that M100 enhancement was instead observed when the pitch of the perceived vowel was shifted (as long as the shift was not too extreme), suggesting that the amplitude of the M100 indexes a comparison between expected and actual stimulation (see Ylinen et al., in press for related findings).

So far we have assumed that forward production and forward comprehension models are taken offline to generate images in self-imagination. But it is also possible that the entire

RUNNING HEAD: Prediction and imagination

process of production (minus overt articulation and overt phonation) is taken offline for the purpose of self-imagination. In this case, not only the forward models but also the implementer (or part of it) would be used to generate images of one's own utterances. This idea is not new; in fact, some version of it corresponds to what many psycholinguists have traditionally assumed about the nature of inner speech. ~~Note that there are at least two contrasting accounts of what inner speech is (Oppenheim & Dell, 2010), but all accounts assume that~~ The idea is that what we experience as inner speech is the outcome of a truncated language production process, and therefore involves the activation of production representations within the implementer (Oppenheim & Dell, 2010). Such production representations are also accessible to the comprehension implementer, at least at the phonological level (Levelt, 1989). However, researchers disagree on whether activation proceeds down to the articulatory level of representation (Corley, Brocklehurst, & Moat, 2011), or whether it stops earlier (Oppenheim & Dell, 2008; MacKay, 1992).

In addition, truncated implementation has been invoked as a component mechanism in the generation of speech-related auditory images, with some studies demonstrating that both implemented production (the "inner voice") and implemented comprehension (the "inner ear") play a role in the creation and categorization of such images (e.g., Reisberg, Smith, Baxter, & Sonenshine, 1989; J.D. Smith, Wilson, & Reisberg, 1995). For example, if either the inner voice or the inner ear was blocked by asking participants to repeat a syllable (articulatory suppression) or by playing irrelevant speech to them (auditory distraction), their ability to categorize images was negatively affected (J.D. Smith et al., 1995). Finally, there may be parallels between inner speech and sub-vocal rehearsal (MacKay, 1992; see also Perrone-Bertolotti, Rapin, Lachaux, Baciú, & Lœvenbruck, 2014), a proposal which is a central to Baddeley's (1986) theory of working memory.

Whereas inner speech (and rehearsal) are not normally considered forms of imagery, in practice relatively similar tasks have been used to investigate inner speech and speech-related auditory imagery. For example, in order to study the generation of auditory images, J.D. Smith et al. (1995) asked participants to interpret string of symbols (e.g., NE1 4 10S)

RUNNING HEAD: Prediction and imagination

that, when pronounced covertly, gave rise to familiar strings of words (e.g., “An-y-one for tenn-is?”). To investigate the nature of inner speech, instead, Wheeldon and Levelt (1995) presented English words to English-Dutch bilinguals over headphones and asked them to mentally translate the words to Dutch.

Interestingly, production mechanisms activated during inner speech may affect concurrent comprehension of individual sounds. Scott and colleagues found that covert rehearsal of a sound affects the way we categorize concurrently presented ambiguous sounds: Perception is “attracted” towards silently rehearsed sounds, particularly when participants are told to move their mouth as they would do when speaking overtly (M. Scott, Yeung, Gick, & Werker, 2013). Scott ~~and colleagues~~et al. attribute these findings to the operation of forward models as discussed above. However, it is possible that internally-generated comprehension representations may be influencing the perception of external sounds, particularly if inner speech is a form of truncated implementation. Therefore Scott et al.’s findings cannot be taken as demonstrating that a simulation process involving forward models is involved in inner speech.

In sum, speech-related auditory images and inner speech might be very similar phenomena, difficult to tease apart experimentally. In any case, we deem it very likely that participants tested in the tasks mentioned above made use of some form of truncated implementation while generating speech images (note that truncated implementation does not imply activation of articulatory representations). It is debatable whether these phenomena ~~can~~ should be considered instances of self-imagination according to our definition of imagination, or rather cases of covert speech preparation, ~~but we acknowledge that the distinction might be more terminological than real.~~

To conclude, we have proposed that self-imagination might sometimes rely on the same mechanisms that allow prediction of one’s own utterances. In doing so, we extended Pickering and Garrod’s (2013) proposal about prediction to imagination, and hypothesized that images of one’s own utterances might be the output of forward production and

RUNNING HEAD: Prediction and imagination

comprehension models. The next section discusses how these ideas can be applied to predicting and imagining others' utterances.

Predicting ~~and imagining~~ other people's utterances using prediction-by-simulation

What happens when I predict somebody else's utterances? It is generally assumed that comprehenders try to recover the communicative intention underlying a speaker's utterances (e.g., Levinson, 2000). In other words, the process of understanding an utterance involves a mapping from acoustic input to the communicative intention that caused the speaker to produce that utterance. There is also consensus among psycholinguists that this mapping is established incrementally: Comprehenders analyse the input bit-by-bit, as soon as it reaches their sensory organs (e.g., Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). Moreover, many scholars believe that comprehenders do not simply analyse the sensory input in a bottom-up fashion and then try to integrate each new input unit into the representation of the speaker's intention they are currently building. Instead, they use the interpretation built so far to generate predictions about upcoming input (e.g., N.J. Smith & Levy, 2013; Altmann & Mirković, 2009).

In Pickering and Garrod's (2013) proposal, comprehenders can generate predictions in two ways. Informally, they can do so by determining what they would do if they were in the speaker's shoes, making use of regularities about their own behaviour in the past and then adjusting for differences between themselves and the speaker. This *prediction-by-simulation* route draws on the forward model mechanisms described above, and requires the comprehender to make reference to (a representation of) the intention of the speaker (Figure 1B). Alternatively, they can make use of regularities in what they have heard in the past (from this speaker or others), just as they would do if predicting an auditory event caused by (non-agentive) objects (e.g., the sound of a wave plus the presence of rocks leads to a prediction of a crashing sound). This *prediction-by-association* route does not depend on intention recognition (Figure 1C). We propose that people can imagine other people's

RUNNING HEAD: Prediction and imagination

utterances by applying either route – leading to *imagination-by-simulation* and *imagination-by-association* (shaded areas in Figure 1B and 1C, respectively). We consider each of these in turn.

Prediction-by-simulation begins with the comprehender deriving expectations about the speaker’s intention (i.e., pre-linguistic expectations), ~~on the basis of contextual or background information~~. Pickering and Garrod (2013) assume that this typically follows from covert imitation of the utterance so far (e.g., *I want a ...*), using this to derive the intention that would underlie this utterance (e.g., expressing a desire), and then deriving the speaker’s likely upcoming intention (i.e., what is the speaker likely to want?). To do this, the comprehender draws on context so that the derived intention corresponds (ideally, at least) to the speaker’s intention (e.g., if the speaker is gazing at a jar full of sweets while saying “I want a”, then it is likely he or she wants wanting – something sweet) rather than the comprehender’s intention (~~underlying what the comprehender would say~~ i.e., what the comprehender would like to have). ~~However, it~~ It is also possible to derive this intention purely from the context (e.g., child pointing towards a jar of sweets) rather than from the context plus the speaker’s words. In either case expectations about upcoming intentions are at least partly based on previously experienced contingencies, in which a given context regularly co-occurred with a given intention.

Then, the comprehender uses intention-to-production and production-to-comprehension mappings (as specified by her own forward production and forward comprehension models) to generate specific linguistic predictions about what the speaker will say; for example, this could involve predicting that the speaker is about to produce the name of some food stuff that begins with /s/. As with self-prediction, a slightly different account of other-prediction by simulation is possible that allows for a direct mapping between intentions and predicted comprehension-based representations, without intermediate production-based representations (at least at some levels). But crucially, prediction-by-simulation involves the comprehender using the speaker’s intention and feeding it into the model(s) she also uses when she is preparing to speak.

RUNNING HEAD: Prediction and imagination

There is converging evidence for prediction of others' utterances via the simulation route. First, articulator-specific motor activation occurs during speech perception, at least under adverse listening conditions (in noise, D'Ausilio, Bufalari, Salmas, Fadiga, 2012; Adank, 2012; when listening to non-native speech, Callan, Jones, Callan, & Akahane-Yamada, 2004), although it is not necessary for speech perception (S.K. Scott, MacGettigan, & Eisner, 2009; Hickok, 2012b). Furthermore, changing the state of the articulators can affect what sounds we actually hear when listening to ambiguous input (Ito, Tiede, & Ostry, 2009), and listening to one syllable can shift the manner of articulation of a concurrently produced syllable in the direction of the manner of articulation of the heard syllable (Yuen, Danis, Brysbaert, & Rastle, 2010).

Second, some findings suggest that production-mechanisms activated during comprehension might support prediction. Thus, Federmeier, Kutas, and Shul²s (2010) found that the magnitude of a late prefrontal positivity induced by plausible but unexpected nouns (which might index prediction updating) correlates with production measures of verbal fluency in older adults. Additionally, rTMS of the right cerebellum delays predictive eye-movements to upcoming linguistic referents (Lesage, Morgan, Olson, Meyer, & Miall, 2012). The cerebellum is known to be implicated in the computation of motor-to-auditory mappings in manual tasks such as finger-tapping (e.g., Knolle, Schröger, & Kotz, 2013) and it might compute similar mappings during speech comprehension; however, direct evidence for the latter hypothesis is lacking.

Imagining other people's utterances using imagination-by-simulation

Having introduced prediction-by-simulation, we now propose that these same predictive mechanisms could be taken offline to imagine others' utterances. A related idea has been put forward by Grush (2004). According to Grush, some perceptual images are generated by running an internal model (in his terminology, an emulator) of the motor process that would generate the "real" perception, and then "taking a perceptual measurement" (p. 388) of the outcome of this model. Our proposal is that sometimes I form images of others'

RUNNING HEAD: Prediction and imagination

utterances by running my own forward production model and then taking the output of the forward production model through the forward comprehension model to generate a comprehension-based image (although sometimes I might “skip” the intermediate step of computing production representations; see also Grush, 2004).

On occasion, I might form the image on the basis of my own forward production and comprehension models, without any adjustments to compensate for differences that might exist between myself and the other person that I am trying to imagine. This might occur whenever I have not had sufficient experience of the target of imagery to learn what adjustments are required. In such cases, the image of another’s utterance would necessarily be formed “from my own perspective”. For example, say I am trying to imagine a child saying *sweet* but I have had little or no experience listening to children saying this word. In that case, I can still imagine the child saying *sweet*, but I would need to rely on my own experience producing this word. My image of the child saying *sweet* would therefore “sound” very much like me saying *sweet*. This situation would then effectively be equivalent to a case of self-imagination. Another example is trying to imagine what it feels like (in terms of somatosensory feedback) for the child to produce the word; almost certainly, I would need to rely heavily on how it feels for me (and this will be substantially different from what it feels like for the child, who has a completely different vocal tract).

But when I imagine somebody with whom I have had extensive interactions, I might be able to correct for the differences between myself and the target of imagery. For example, if I have been listening to the particular child I am trying to imagine for a long time, I could use adjusted forward production and comprehension models to generate an image “in the child’s voice”. In such cases, When the production route to imagery is not necessary, but possible. Moreover, if it is used, it should produce images that are faithful to the target of imagery. Crucially, such images would be generated on the basis of the communicative intention I attribute to the child. Regardless of whether the output of imagination by simulation is accurate or not, we propose that imagining another’s utterances proceeds via simulation when the target of imagination is perceived as being sufficiently similar to oneself.

RUNNING HEAD: Prediction and imagination

[because in such cases simulation is more likely to be accurate](#) (see Pickering & Garrod, 2013, p. 346 for ~~the same~~ [an equivalent](#) proposal ~~regarding about~~ prediction-by-simulation).

An alternative route to predicting and imagining others' utterances: association.

[According to Pickering and Garrod \(2013\)](#), ~~t~~he second way in which I can predict others' utterance is via prediction-by-association. This predictive mechanism exploits co-occurrences in the input to the process of comprehension and makes no reference to intentions. For example, I can predict the sound of leaves rustling based on feeling the wind blow, because these two perceptual events have co-occurred frequently in the past, but I most likely do not make this prediction on the basis on a representation of the wind's intention to rustle the leaves, nor simulate the way in which the wind moves the branches or the movement of the branches produces the rustling sound.³ In our example, I predict that I will hear the word *sweet* because in my previous experience of children in similar contexts (or of this particular child in this particular context) the words *I want a* have been followed by *sweet* more often than they have been followed by *swede*.

Similarly, we propose that people can imagine others' utterances via this association route as well. For example, I can imagine a child saying *sweet* (rather than *swede*) via the association route, provided I have experienced many instances of the situation in which a child is present and the word *sweet* is uttered. If I have heard this particular child saying *sweet* many times in the past, I could even retrieve an accurate memory representation of what the word sounds like in the child's voice. Crucially, this would not require any simulation of the child's production system. Actually, it would not require attributing the child an intention to speak, as the same route to imagery is available for perceptual events caused by inanimate entities to which we do not attribute intentions (as the wind in our previous example).

It is possible that imagination-by-association is the default way of imagining others' utterances. Previous literature has indeed suggested there is a difference between imagining oneself producing language and imagining hearing another person doing so (Tian & Poeppel, 2010; 2013). If I imagine myself articulating a syllable, subsequent early auditory responses

RUNNING HEAD: Prediction and imagination

(M200) to the same syllable are enhanced, just as when I have actually articulated the syllable. However, if I imagine another person (the experimenter) saying the syllable, the M200 is suppressed when the same syllable is presented again, similarly to when I have actually heard the syllable spoken in another’s voice (Tian & Poeppel, 2013).

One interpretation of this finding is that when I imagine hearing a syllable in somebody else’s voice, I do not compute intention-to-production and production-to-comprehension mappings, but I use prediction-by-association. On a different interpretation, people might imagine hearing another person using the simulation route, just as they do when they imagine themselves speaking. However, they might form much more detailed images when imagining themselves compared to when they imagine others. As a consequence, they might also be more sensitive to mismatches between their simulated images and the target sound in the self-imagination compared to the other-imagination condition. The enhanced M200 response could be the neural correlate of this enhanced sensitivity to mismatches. One way of differentiating between these two explanations, and determining whether people can use imagination-by-simulation to imagine others’ utterances, would be to familiarize the participants extensively with the target of their imagination and verify whether in this case a pattern of M200 enhancement also occurs in the other-imagination condition.

As an aside, we also note that it is uncertain whether one could use the association route to imagine one’s own utterances (a possibility we have not taken into account in section 2). For example, I could imagine listening to my voice from “outside”, as might be the case when my voice is played back to me after having been recorded. Of course my voice would sound somewhat “alien” in this case, to the extent that I might not recognize it as mine when it is played back to me. It is possible that this might ~~facilitate using~~ encourage the use of association-based imagery instead of production-based imagery in such cases. But if this is the case, then this situation effectively reduces to an example of imagination of others’ utterances.

To summarize, we have proposed that one way we can imagine both our own and others’ utterances is via simulation, which is based on intentions. This is the same mechanism

RUNNING HEAD: Prediction and imagination

that we use to predict aspects of the production and comprehension of our [own](#) and others' utterances (when we prepare to say something or try to ~~anticipate-predict~~ what somebody else is ~~about-going~~ to say using the simulation route). But in addition, the association route is also available to predict and imagine others' utterances.

In conclusion, we propose two distinctions, which are in principle orthogonal to one another: the distinction between imagery of oneself versus another; and the distinction between the association route and the simulation route to imagery. To our knowledge, nobody has tested whether imagining hearing one's own voice is the same as imagining articulation, or whether imagining hearing somebody else's voice is the same as imagining them articulating. However, it is likely that imagining one's own utterances proceeds by default via the simulation route, and that imagining another's utterances proceeds via simulation route only under certain conditions (i.e., when I perceive the target of imagination as being sufficiently similar to myself). See Figure 1D for a summary of our proposal.

Imagining and predicting others in joint tasks

So far we have looked at prediction and imagination of utterances produced by isolated speakers. But clearly most of language production and comprehension takes place in conversation, and therefore most of prediction takes place in conversation. Moreover, Pickering and Garrod (2013) suggested that prediction-by-simulation could explain smooth conversational exchanges, supporting both timely turn-taking and rapid preparation of appropriate responses to the interlocutor's contributions. In a nutshell, if predicting another's utterances activates the forward production and comprehension models, the task of comprehending what the other is saying will facilitate *subsequent* production (see also Pickering and Garrod, 2004). Conversely, since comprehenders are also speakers in conversation, [and they use production mechanisms, including forward models, when speaking,](#) the use of forward production and comprehension models for prediction-by-simulation of others' utterances could be primed [in conversation, compared to listening to a](#)

RUNNING HEAD: Prediction and imagination

[monologue](#) (Pickering & Garrod, 2013). Prediction-by-simulation is one mechanism that could allow a tight integration of self and other during conversational interactions.

In recent work, we tested whether people represent others' utterances similarly to their own utterances using joint language tasks (Gambi & Pickering, 2013). Such tasks are very different from conversational interactions, as we elicit utterances using pictures and regulate turn-taking via task instructions. Nevertheless, they allow one to investigate whether people simulate one another's utterances when placed in a situation that provides some of the basic ingredients of dialogue (e.g., a joint goal, a turn-taking system). Our studies are informed by the extensive literature on non-linguistic joint tasks. This body of research has showed that when participants are asked to perform a task together, they represent their partner's task at some level (Knoblich, Butterfill, & Sebanz, 2011), and possibly even simulate some aspects of the partner's action. For example, Loehr, Kourtis, Vesper, Sebanz, and Knoblich (2013) used electroencephalography to show that musicians playing a duet respond to externally-induced "mistakes" in the same way regardless of whether their own or their partner's portion of the duet is affected. This study shows that people monitor others' actions as if they were their own, which in turn suggests that they construct internal simulations of what it should be like to perform those actions.

In the language domain, Baus et al. (2014) found that speakers represent how difficult an act of naming will be for their partner. In this study, a participant and a confederate took turns in naming pictures with high or low frequency names. A broadly-distributed P300 amplitude difference between low and high frequency words occurred when the participant was not responding but the confederate was preparing to name a picture, whereas no such effect occurred when nobody was preparing to name the picture (but the confederate was still present). This suggests that participants simulated their partner's act of naming using the same intention-to-production mappings that they used when preparing to name the pictures themselves.

In a behavioural study, we asked participants to take turns naming pictures with another person (Gambi, Cop, & Pickering, 2014in press). On the critical trials, the initially

RUNNING HEAD: Prediction and imagination

presented picture changed into a different picture, and participants were instructed to stop speaking as quickly as possible. After stopping, one group was told that the same person who stopped had to name the new picture, another group was told that the other person should name the new picture instead, and a third group had to ignore the new picture. People who had to stop and then speak were less likely to stop mid-word than people who simply had to stop. Importantly, people who had to stop and let their partner speak were also less likely to stop mid-word than people who just stopped, although they were still more likely to do so than people who stopped and then spoke again. This shows that predicting that another person will speak affects the process of stopping speech similarly to predicting oneself (although to a lesser extent) and suggests that self- and other-prediction ~~might~~ make use of ~~some~~ a common mechanism.

It would be interesting to test whether purely imagining that another person will name the new picture would have a comparable effect. There is some evidence that people represent their partner's task even when they are seated in different rooms and therefore cannot perceive their partner's actions, but can only imagine those actions based on the instructions received by their partners (e.g., Atmaca, Sebanz, & Knoblich, 2011; Exp. 3). Recently, we showed that people cannot help but imagining another person's utterances, even if they can never hear those utterances (Gambi, Van de Cavey, & Pickering, ~~2014~~2015). Specifically, we asked speakers to name pictures while their partner either named pictures, remained silent, or categorized the pictures as belonging to the same or to different semantic categories (before each naming trial the participant's and the partner's names were displayed to both with their respective instructions, so that they knew what their partner would be doing). Speakers took longer to begin naming when their partner was also naming. This suggests that imagining that another is speaking relies on some of the same mechanisms involved in preparation for speech. This shared mechanism could be a forward production model that maps from the partner's intention to name the pictures (an intention that participants can derive from the instructions on the screen) to the state of the production system once that intention has been

RUNNING HEAD: Prediction and imagination

realized. In other words, the study suggests that imagining another’s utterances can be based on the simulation route, at least when the other person is their partner in a joint task.

Another interpretation of this study is that people slowed down because imagining their partner say something at the same time as them caused them to prepare for a “choral act of production”. Or, to put it slightly differently, they might have imagined that they were not speaking on their own, but rather speaking at the same time as another. We know that in actual choral productions people tend to slow down their speech rate (Cummins, 2003), ~~presumably-perhaps~~ because participants try in this way to facilitate coordination with the other speaker. or because there are using some resources to simulate their partner, and have less resources available for production.

More generally, one might ask what it means to imagine joint language uses as opposed to individual language use; that is, what does it mean to imagine a dialogue, or an argument? Do people simply alternate between self-imagination, which has to rely on simulation, and other-imagination, which usually takes the association route? Or would they be particularly likely to use the simulation route for other-imagination as well (just as they would be more likely to use prediction-by-simulation if they were having a real dialogue)? ~~This latter possibility has the advantage of not requiring switching between routes. We think the second alternative might be correct. Interestingly, the~~ The way in which participants coordinate their movements in a joint jumping task obeys the same rules whether the jumps are actually executed (Vesper, Van der Wel, Knoblich, & Sebanz, 2013) or just imagined (Vesper, Sebanz, & Knoblich, 2014). But we are not aware of any study that looked at the imagination of dialogue, so this question remains open. In sum, we propose that people tend to use the simulation route in other-imagination when they are imagining the utterances produced by their partner in a joint task, particularly when these imagined utterances are part of an imagined interaction (Figure 1D).

Conclusion

RUNNING HEAD: Prediction and imagination

In conclusion, we proposed that imagination of utterances is a form of offline prediction of utterances. We argued that imagining one's own utterances differs from imagining others' utterances. While I imagine myself by simulating production, I ~~usually~~ often imagine others using associations between comprehension-based representations. However, ~~in certain situations~~ on other occasions (when I perceive the other as being similar to myself, and when the other is my partner in a joint activity) I might imagine others like I imagine myself, by building a simulation from intention to production and comprehension representations.

RUNNING HEAD: Prediction and imagination

References

- Adank, P. (2012). The neural bases of difficult speech comprehension and speech production: ~~two~~ Two activation likelihood estimation (ALE) meta-analyses. *Brain and Language*, 122(1), 42-54.
- Altmann, G., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73(3), 247-264.
- Altmann, G., & Mirković, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science*, 33(4), 583-609.
- Atkinson, J. R. (2006). The perceptual characteristics of voice-hallucinations in deaf people: insights into the nature of subvocal thought and sensory feedback loops. *Schizophrenia Bulletin*, 32(4), 701-708.
- Atmaca, S., Sebanz, N., & Knoblich, G. (2011). The joint flanker effect: Sharing tasks with real and imagined co-actors. *Experimental Brain Research*, 211(3-4), 371-385.
- Baddeley, A. D. (1986). *Working memory*. Oxford: Clarendon Press.
- Baus, C., Sebanz, N., de la Fuente, V., Branzi, F. M., Martin, C., & Costa, A. (2014). On predicting others' words: Electrophysiological evidence of prediction in speech production. *Cognition*, 133(2), 395-407.
- Behroozmand, R., & Larson, C. R. (2011). Error-dependent modulation of speech-induced auditory suppression for pitch-shifted voice feedback. *BMC ~~neuroscience~~ Neuroscience*, 12(1), 54.
- 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(3), 1182-1194.
- 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), 162:175.
- Cummins, F. (2003). Practice and performance in speech produced synchronously. *Journal of*

RUNNING HEAD: Prediction and imagination

Phonetics, 31(2), 139-148.

D'Ausilio, A., Bufalari, I., Salmas, P., & Fadiga, L. (2012). The role of the motor system in discriminating normal and degraded speech sounds. *Cortex*, 48(7), 882-887.

Emmorey, K., Bosworth, R., & Kraljic, T. (2009). Visual feedback and self-monitoring of sign language. *Journal of Memory and Language*, 61(3), 398-411.

Federmeier, K. D., Kutas, M., & Schul, R. (2010). Age-related and individual differences in the use of prediction during language comprehension. *Brain and Language*, 115(3), 149-161.

Gambi, C., & Pickering, M. J. (2013). Talking to each other and talking together: Joint language tasks and degrees of interactivity. *Behavioral and Brain Sciences*, 36(04), 423-424.

[Gambi, C., & Pickering, M. J. \(forthcoming\). Models linking production and comprehension.](#)

[In Fernández, E., and Cairns, H. \(Eds.\). *Handbook of Psycholinguistics*. Wiley-Blackwell.](#)

Gambi, C., Cop, U., & Pickering, M. J. (2014in press). How do speakers coordinate?

Evidence for prediction in a joint word-replacement task. *Cortex*,

doi:10.1016/j.cortex.2014.09.009

Gambi, C., Van de Cavey, J., & Pickering, M.J. (20142015). Interference in joint picture naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(1), 1-21,

doi: 10.1037/a0037438

Goldman, A. I. (2006). *Simulating minds: The philosophy, psychology, and neuroscience of mindreading*. Oxford: Oxford University Press.

Grush, Riek. "The emulation theory of representation: ~~motor~~Motor control, imagery, and perception." *Behavioral and Brain Sciences*, 27(3), 377-396.

Heinks-Maldonado, T. H., Nagarajan, S. S., & Houde, J. F. (2006). Magnetoencephalographic evidence for a precise forward model in speech production. *Neuroreport*, 17(13), 1375-1379.

Hickok, G. (2012a). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*, 13(2), 135-145.

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(6), 393-402.

RUNNING HEAD: Prediction and imagination

- 1
- 2
- 3 Hurley, S. (2008). Understanding Simulation. *Philosophy and Phenomenological Research*,
- 4
- 5 77(3), 755-774.
- 6
- 7 Ito, T., Tiede, M., & Ostry, D. J. (2009). Somatosensory function in speech perception.
- 8
- 9 *Proceedings of the National Academy of Sciences*, 106(4), 1245-1248.
- 10
- 11 Kempen, G. (2014). Prolegomena to a neurocomputational architecture for human
- 12
- 13 grammatical encoding and decoding. *Neuroinformatics*, 12(1), 111-142.
- 14
- 15 Knolle, F., Schröger, E., & Kotz, S. A. (2013). Cerebellar contribution to the prediction of
- 16
- 17 self-initiated sounds. *Cortex*, 49(9), 2449-2461.
- 18
- 19 Kutas, M., DeLong, K. A., & Smith, N. J. (2011). A look around at what lies ahead:
- 20
- 21 Prediction and predictability in language processing. In M. Bar (Ed.), *Predictions in the*
- 22
- 23 *brain: Using our past to generate a future* (pp. 190-207). Oxford: Oxford University Press.
- 24
- 25 Lesage, E., Morgan, B. E., Olson, A. C., Meyer, A. S., & Miall, R. C. (2012). Cerebellar
- 26
- 27 rTMS disrupts predictive language processing. *Current Biology*, 22 (18), R794-R795.
- 28
- 29 Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT
- 30
- 31 Press.
- 32
- 33 Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational*
- 34
- 35 *implicature*. Cambridge, MA: MIT press.
- 36
- 37 Loehr, J. D., Kourtis, D., Vesper, C., Sebanz, N., & Knoblich, G. (2013). Monitoring
- 38
- 39 individual and joint action outcomes in duet music performance. *Journal of Cognitive*
- 40
- 41 *Neuroscience*, 25(7), 1049-1061.
- 42
- 43 MacKay, D.G. (1992). Constraints on theories of inner speech. In Reisberg, D. (Ed.),
- 44
- 45 *Auditory imagery*, (pp. 121-149). Hillsdale, NJ, ~~England: Lawrence-Erlbaum-Associates.~~
- 46
- 47 Niziolek, C. A., Nagarajan, S. S., & Houde, J. F. (2013). What does motor efference copy
- 48
- 49 represent? Evidence from speech production. *The Journal of Neuroscience*, 33(41), 16110-
- 50
- 51 16116.
- 52
- 53 [Nozari, N., Dell, G. S., & Schwartz, M. F. \(2011\). Is comprehension necessary for error](#)
- 54
- 55 [detection? A conflict-based account of monitoring in speech production. *Cognitive*](#)
- 56
- 57 [Psychology](#), 63(1), 1-33.
- 58
- 59
- 60

RUNNING HEAD: Prediction and imagination

- Oppenheim, G. M., & Dell, G. S. (2008). Inner speech slips exhibit lexical bias, but not the phonemic similarity effect. *Cognition*, 106(1), 528-537.
- Oppenheim, G. M., & Dell, G. S. (2010). Motor movement matters: The flexible abstractness of inner speech. *Memory & Cognition*, 38(8), 1147-1160.
- Perrone-Bertolotti, M., Rapin, L., Lachaux, J. P., Baci, M., & Lœvenbruck, H. (2014). What is that little voice inside my head? Inner speech phenomenology, its role in cognitive performance, and its relation to self-monitoring. *Behavioural Brain Research*, 261, 220-239.
- Pezzulo, G. (2011). Grounding procedural and declarative knowledge in sensorimotor anticipation. *Mind & Language*, 26(1), 78-114.
- Pickering, M. J., & Clark, A. (2014). Getting ahead: Forward models and their place in cognitive architecture. *Trends in Cognitive Sciences*, 18(9), 451-456.
- Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain sciences*, 27(2), 169-190.
- Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, 36(4), 329-392.
- Pickering, M. J., & Garrod, S. (2014). Self-, other-, and joint monitoring using forward models. *Frontiers in Human Neuroscience*, 8, doi: [10.3389/fnhum.2014.00132](https://doi.org/10.3389/fnhum.2014.00132).
- Reisberg, D., Smith, J. D., Baxter, D. A., & Sonenshine, M. (1989). "Enacted" auditory images are ambiguous; "pure" auditory images are not. *The Quarterly Journal of Experimental Psychology*, 41(3), 619-641.
- Scott, M., Yeung, H. H., Gick, B., & Werker, J. F. (2013). Inner speech captures the perception of external speech. *The Journal of the Acoustical Society of America*, 133(4), EL286-EL292.
- Scott, S. K., 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(4), 295-302.
- Smith, J. D., Wilson, M., & Reisberg, D. (1995). The role of subvocalization in auditory imagery. *Neuropsychologia*, 33(11), 1433-1454.

RUNNING HEAD: Prediction and imagination

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, 128(3), 302-319.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268(5217), 1632-1634.

Thomas, N. J. (1999). Are theories of imagery theories of imagination?: An active perception approach to conscious mental content. *Cognitive Science*, 23(2), 207-245.

Tian, X., & Poeppel, D. (2010). Mental imagery of speech and movement implicates the dynamics of internal forward models. *Frontiers in Psychology*, 1, DOI: 10.3389/fpsyg.2010.00166.

Tian, X., & Poeppel, D. (2012). Mental imagery of speech: Linking motor and perceptual systems through internal simulation and estimation. *Frontiers in Human Neuroscience*, 6, doi: 10.3389/fnhum.2012.00314

Tian, X., & Poeppel, D. (2013). The effect of imagination on stimulation: the functional specificity of efference copies in speech processing. *Journal of Cognitive Neuroscience*, 25(7), 1020-1036.

[Tian, X., & Poeppel, D. \(2015\). Dynamics of self-monitoring and error detection in speech production: evidence from mental imagery MEG. *Journal of Cognitive Neuroscience*. 27\(2\), 352-364.](#)

Tourville, J. A., & Guenther, F. H. (2011). The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes*, 26(7), 952-981.

Van Berkum, J. J., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: ~~evidence~~ Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(3), 443.

Vesper, C., Knoblich, G., & Sebanz, N. (2014). Our actions in my mind: Motor imagery of joint action. *Neuropsychologia*, 55, 115-121.

RUNNING HEAD: Prediction and imagination

Vesper, C., Van der Wel, R. P. R. D., Knoblich, G., & Sebanz, N. (2013). Are you ready to jump? Predictive mechanisms in interpersonal coordination. *Journal of Experimental Psychology: Human Perception and Performance*, 39(1), 48-61.

Wheeldon, L. R., & Levelt, W. J. (1995). Monitoring the time course of phonological encoding. *Journal of Memory and Language*, 34(3), 311-334.

[Ylinen, S., Nora, A., Leminen, A., Hakala, T., Huotilainen, M., Shtyrov, Y. et al. \(in press\). Two distinct auditory-motor circuits for monitoring speech production as revealed by content-specific suppression of auditory cortex. *Cerebral Cortex*.
<http://dx.doi.org/10.1093/cercor/bht351>](#)

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(2), 592-597.

RUNNING HEAD: Prediction and imagination

Figure 1. Diagrams showing (A) prediction by simulation in an instance of language production and in self-imagination; (B) prediction by simulation in an instance of language comprehension, and in other-imagination; (C) prediction by association in an instance of language comprehension, and in other-imagination. Processes extracted for offline use in imagination are shaded in grey. Figures in (A) and (B) are simplified versions of Pickering and Garrod’s (2013) Figures 5 and 6. (D) Summary of our proposal regarding types of imagination.

RUNNING HEAD: Prediction and imagination

Footnote 1. We use imagination and imagery interchangeably. Our usage contrasts with accounts in which imagination includes forms of creative thinking and imagery refers only to quasi-perceptual experiences (e.g., Thomas, 1999).

Footnote 2. There is no direct parallel between Tian and Poeppel's (2012) somatosensory forward model and Pickering and Garrod's (2013) forward production model; crucially, Pickering and Garrod's forward production model maps the speaker's intention to production representations, which are not sensory in nature.

Footnote 3. It may be possible to partly anthropomorphise the behaviour of non-intentional objects by making use of simulation, especially if they mimic intentional agents (e.g., a speech dialogue system that interacts in human-like manner).

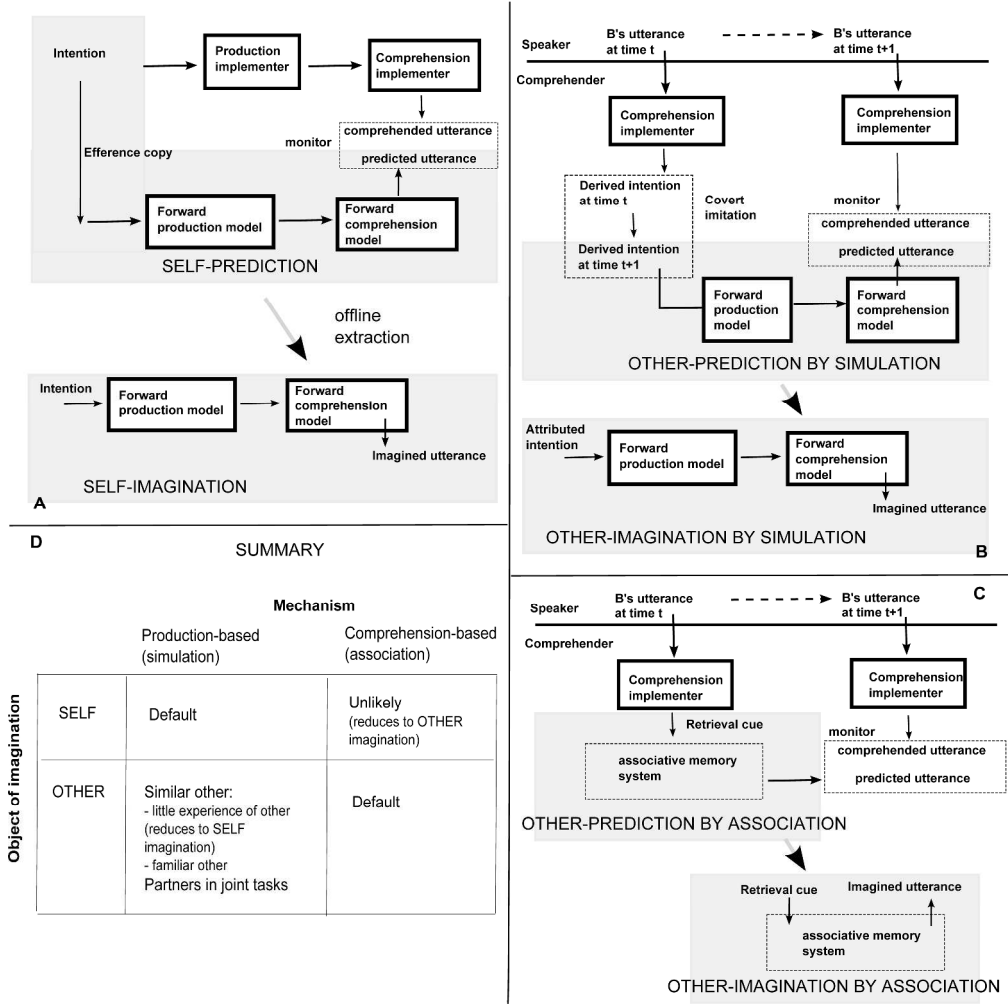


Figure 1
456x457mm (300 x 300 DPI)