

Joint Action, Interactive Alignment, and Dialog

Simon Garrod,^a Martin J. Pickering^b

^a*Department of Psychology, University of Glasgow*

^b*Department of Psychology, University of Edinburgh*

Received 17 July 2008; received in revised form 23 October 2008; accepted 10 November 2008

Abstract

Dialog is a joint action at different levels. At the highest level, the goal of interlocutors is to align their mental representations. This emerges from joint activity at lower levels, both concerned with linguistic decisions (e.g., choice of words) and nonlinguistic processes (e.g., alignment of posture or speech rate). Because of the high-level goal, the interlocutors are particularly concerned with close coupling at these lower levels. As we illustrate with examples, this means that imitation and entrainment are particularly pronounced during interactive communication. We then argue that the mechanisms underlying such processes involve covert imitation of interlocutors' communicative behavior, leading to emulation of their expected behavior. In other words, communication provides a very good example of predictive emulation, in a way that leads to successful joint activity.

Keywords: Interactive alignment; Dialog; Emulation; Prediction; Joint action

1. Introduction

Psychologists have recently begun to address the mechanisms underlying the integration of perception, action, and cognition across individuals. It is clear that an understanding of linguistic communication requires the study of interacting individuals, as many sociological investigations of language show (e.g., Goffman, 1981; Sacks, Schegloff, & Jefferson, 1974). But traditional mechanistic psycholinguistics focuses on individual acts of production or comprehension and seeks to determine the nature of the mental representations and processes used by the isolated individual (see Traxler & Gernsbacher, 2006). So how can we understand the mechanisms of linguistic communication in terms of joint action, in which the action depends on the integrated behavior of more than one individual?

Correspondence should be sent to Simon Garrod, Department of Psychology, University of Glasgow, 56 Hillhead Street, Glasgow G12 8QB, UK. E-mail: simon@psy.gla.ac.uk

In this paper, we argue that mechanisms of communication used in dialog depend on joint action. When people interact, their representations become more aligned at many different levels, from basic motor programs to high-level aspects of meaning. We argue that such processes of alignment are almost certainly enhanced by having (or perhaps sharing) the goal of communicating with each other. But at the same time, these influences are largely automatic (so that interactants are typically unaware of the alignment processes). We argue that both linguistic and nonlinguistic processes make use of emulation and prediction to drive joint action. These processes are enhanced by the existence of the communicative goal, in a way that makes dialog a particularly well-integrated form of joint action.

2. The success of dialog

To appreciate why dialog can be seen as a form of joint action, let us consider the following transcript from a cooperative game in which participants need to establish where their partner is located in a maze (Garrod & Anderson, 1987).

1. A: You know the extreme right, there's one box.
2. B: Yeah right, the extreme right it's sticking out like a sore thumb.
3. A: That's where I am.
4. B: It's like a right indicator.
5. A: Yes, and where are you?
6. B: Well I'm er: that right indicator you've got.
7. A: Yes.
8. B: The right indicator above that.
9. A: Yes.
10. B: Now if you go along there. You know where the right indicator above yours is?
11. A: Yes.
12. B: If you go along to the left: I'm in that box which is like: one, two boxes down O.K.?

Although many of the individual contributions may be difficult to understand in isolation (e.g., 1, 2, 8), they make sense as part of the whole interaction. This interdependence makes it clear why they should be interpreted as part of a joint action.

In fact, dialog is a remarkably successful form of joint action, given its complexity. As pointed out by Garrod and Pickering (2004), there is a sense in which dialog ought to be extremely difficult, in comparison to monolog. Interlocutors have to produce responses on the fly (e.g., 6), to comprehend elliptical utterances (e.g., 8), to repeatedly determine who to speak to and precisely when to speak, to comprehend and produce at the same time (when receiving feedback), and of course to task-switch constantly between comprehension and production. But it is also clear that interlocutors do not typically find dialog difficult. Thus, it occurs in the earliest stages of caregiver-child interaction, whereas giving a speech or even listening to one is clearly a difficult-to-acquire skill. Evidence for the ease of dialog

comes from studies that compare narratives in which the audience provides good feedback (which is a form of dialog) with narratives in which the audience is largely silent (which is therefore similar to monolog). For instance, Bavelas, Coates, and Johnson (2000) told an addressee to count a narrator's use of the letter "t," an activity that greatly interferes with feedback. The narrator's description became much less informative and its ending was more abrupt, repetitive, "choppy," and inappropriate. Hence, the addressee's feedback contributed to the quality of the story (see also Kraut, Lewis, & Swezey, 1982). Even though the narrator has more to do when comprehending feedback as well as story-telling, the dialog is easier and more successful than the (near) monolog.

An explanation of why dialog is successful comes from the interactive-alignment account (Pickering & Garrod, 2004). This account assumes that communication is successful to the extent that communicators come to understand relevant aspects of the world in the same way as each other. In other words, they *align* their representation of the situation under discussion (see also Brown-Schmidt & Tanenhaus, 2008). Alignment is typically achieved (to an extent that makes people believe that they generally understand each other), because people start off at a very good point. They communicate with other people who are largely similar to themselves, both because they process language in similar ways and because they share much relevant background knowledge. This means that they can, in principle, use knowledge about themselves to understand and, in particular, predict their interlocutor.

Of course, it is possible to use such knowledge during monolog as well as dialog. In comprehending monolog, the listener (or reader) can use self-knowledge to determine the speaker's (or writer's) meaning. For instance, when a listener makes an inference about how two sentences are connected, the result is likely to correspond to what the speaker meant as well. But there are many advantages present in a dialog that do not occur in monolog. For example, if the addressee does not understand the speaker at a particular point, he can simply ask her, or provide some other indication (such as *eh?* or a confused expression) that he does not understand. More important, such interaction changes the speaker's representations as well as the addressee's in such a way that the extent of their alignment increases. Therefore, they are more likely to understand each other, to draw the same inferences, and so on. These advantages occur because dialog is a form of joint action. But what exactly does this mean?

3. Dialog as joint action

Most discussions of joint action are concerned with situations in which "two or more individuals coordinate their actions in space and time to bring about a change in the environment" (Sebanz, Bekkering, & Knoblich, 2006, p. 70). Sebanz et al. argue that such joint actions require merging your own action plans with those of your partner, in a way that leads to shared representations. To engage in joint actions requires being able to predict others' actions and integrate the predicted effects of your own and others' actions. For example, to succeed in ballroom dancing it is not sufficient to simply recognize your

partner's movements; you also need to be able to predict them in advance. Otherwise, you would not be able to coordinate the timing of your actions with your partner's.

According to Sebanz et al.'s (2006) definition, joint actions are intentional (or goal-directed)—individuals coordinate for a particular purpose. However, we argue that there are other forms of joint action that are incidental, in that there is no intention to coordinate. For example, Richardson, Marsh, Isenhower, Goodman, and Schmidt (2007b) found that two individuals seated side by side in rocking chairs tended to rock in synchrony with each other. This occurred even when they were instructed to rock at their own pace and had chairs with different natural rocking frequencies. Similar observations have been made about mimicry of incidental actions (e.g., head scratching or foot tapping) during social interaction (Chartrand & Bargh, 1999). Thus, joint actions can be intentional (or goal-directed) or unintentional (or incidental).

Dialog *as a whole* constitutes a form of intentional joint action. The interlocutors have the goal of communicating and realize that their partners also have this goal (Clark, 1996). Of course, people use communication to do many different things. Instructing someone about repairing a car, giving a history lesson, and holding a casual conversation about the weather are examples of different kinds of communication. In fact, we argue that interlocutors share the goal of aligning their representations of the situation under discussion, whether it may be concerned with car maintenance, historical facts, or today's weather (see Garrod & Pickering, 2009). Even when two people are arguing, they still share this goal. It is always the case that interlocutors seek to make their contributions intelligible, for example, by ensuring that they and their partners interpret expressions as referring to the same entities.

Notice that our use of joint action does not require there to be a change in the environment beyond the behavior itself, in contrast to Sebanz et al.'s (2006) definition. So dialog is a joint action even if the interlocutors only interact verbally and the dialog has no further consequences. (This use of joint action corresponds to Clark, 1996.)

Thus, dialog is a form of joint action in which interlocutors have the goal of aligning their understanding of the situation (whether they agree about every detail or not). Pickering and Garrod (2004) argue that interlocutors achieve such alignment by aligning at many other linguistic levels, concerned with words, grammar, and sound, and that such alignment leads to alignment of situation models. We now show how this can come about.

4. Alignment arises from joint actions at many levels

Dialog, like other complex actions, can be decomposed into many separate components. We can therefore see it as involving many separate joint actions (e.g., question–answer pairs such as 5–6 above). At each point in the interaction, each interlocutor makes a series of choices about what to say and how to say it. Should such choices be seen as intentional (and have meaning in the sense of Grice, 1957), simply because the dialog as a whole is intentional? This appears to be the position advocated by Clark (1996), who analyzes dialog as involving many layers of activity. In particular, he argues that speakers intentionally convey meaning at each of these layers, for example, using disfluencies to inform addressees that

they are having difficulty (Clark & Fox Tree, 2002; Fox Tree & Clark, 1997). In contrast, we argue that some aspects of dialog do not depend on intentional activity and should be seen as reflecting automatic processes.

Pickering and Garrod (2004) proposed that the most important way in which alignment occurs is via a process of automatic (nonconscious and effortless) *imitation* at different linguistic levels. For example, when describing pictures of objects to each other, interlocutors use the same expressions to refer to them (e.g., Brennan & Clark, 1996; see also Garrod & Anderson, 1987). Similarly, they tend to repeat each other's choice of grammatical constructions to describe pictures of events (Branigan, Pickering, & Cleland, 2000). If the confederate described a card as *The chef giving the jug to the swimmer*, the participant would tend to describe the next card as *The cowboy handing the banana to the burglar*; but if the confederate said *The chef giving the swimmer the jug*, the participant would then tend to say *The cowboy handing the burglar the banana*. Such effects can be extremely strong and occur for other grammatical constructions (Cleland & Pickering, 2003) and even between languages in bilinguals (Hartsuiker, Pickering, & Veltkamp, 2004).

Pickering and Garrod (2004) argued that such effects occur because interlocutors use *common coding* across production and comprehension, as may be the case for action and perception more generally (e.g., Prinz, 1990). Thus, activating a representation of a word or grammatical construction during comprehension enhances its activation during production. Thus, alignment in dialog occurs automatically, without extensive negotiation between interlocutors or modeling of each other's mental states.

In itself, this account would lead to imitation at different levels of representation, but something more is needed for alignment to "percolate" between levels. In fact, Pickering and Garrod (2004) point out that alignment at one level enhances alignment at other levels. For example, grammatical alignment is enhanced by repetition of words, with participants being even more likely to say *The cowboy handing the banana to the burglar* after hearing *The chef handing the jug to the swimmer* than after *The chef giving the jug to the swimmer* (Branigan et al., 2000). Indeed, grammatical alignment is enhanced for words that are semantically related, with the tendency to say *the sheep that's red* (rather than *the red sheep*) being greater following *the goat that's red* than following *the knife that's red* (Cleland & Pickering, 2003). In this case, semantic alignment enhances grammatical alignment. In the same way, alignment of words leads to alignment of situation models—people who describe things the same way tend to think about them in the same way too (Markman & Makin, 1998). These processes of alignment are not therefore primarily intentional, even though the interlocutors have the goal of aligning their understanding.

There is considerable evidence interlocutors imitate each other at many nonlinguistic levels. For example, interlocutors tend to adopt the same postures as each other (Shockley, Santana, & Fowler, 2003) and to laugh or yawn together (Hatfield, Cacioppo, & Rapson, 1994). This clearly means that interlocutors construct aligned nonlinguistic representations. Just as linguistic alignment at one level can enhance alignment at other levels, so nonlinguistic alignment can also enhance alignment. For example, addressees who align their gaze with speakers tend to align their interpretations with the speaker as well (Richardson & Dale, 2005). Thus, highlighting pictures as they are fixated by the speaker enhances

comprehension. It may also be the case that nonlinguistic alignment indirectly leads to alignment of situation models. For example, people who align on a particular mood are more likely to use the same or semantically similar expressions (Hatfield et al., 1994), and this is likely to lead to alignment of situation models.¹

In summary, the joint action of communication occurs when interlocutors align their situation models. This is the end result of alignment at many other levels, some of which are linguistic and some of which are nonlinguistic. Interlocutors intentionally communicate, in that they set out to align their situation models. But the extent to which alignment at other levels is intentional is likely to vary greatly, and it can be seen as being on a cline. We now turn to the question of how such processes could be realized within interlocutors' minds.

5. Interactive alignment involves emulation

As we have noted, an important aspect of many joint actions is that they require not just perceiving (or comprehending) a partner's actions but also predicting those actions. Failure to predict a ballroom partner's steps will lead to the lack of coordination (if not complete collapse). We argue that the same is true of dialog and that interactive alignment comes about in part through emulation and prediction.

There is considerable evidence that language comprehension (both in dialog and monolog) involves prediction. For example, readers and listeners predict high cloze up-coming words before encountering them. Van Berkum, Brown, Zwitserlood, Kooijman, and Hagoort (2005) had participants listen to or read predictable (or "High-Cloze") texts in Dutch such as the following:

13. *De inbreker had geen enkele moeite de geheime familiekluis te vinden.* [The burglar had no trouble locating the secret family safe.]
- (a) *Deze bevond zich natuurlijk achter een groot_{neuter} maar onopvallend schilderij_{neuter}.* [Of course, it was situated behind a big_{neuter} but unobtrusive painting_{neuter}.] (predictable)
- (b) *Deze bevond zich natuurlijk achter een grote_{common} maar onopvallende boekenkast_{common}.* [Of course, it was situated behind a big_{common} but unobtrusive bookcase_{common}.] (not predictable)

The adjective *groot* in (13a) agrees in gender with the predictable noun *schilderij*, whereas the adjective *grote* in (13b) does not. If participants predict the upcoming noun (and its gender), they should be disrupted when the adjective is incongruent. In accord with this claim, Van Berkum et al. found that participants were disrupted when they encountered *grote* in comparison to *groot*, for spoken language comprehension (using even-related potentials) and written language comprehension (using self-paced reading). Comparable results occurred with article gender in Spanish (Wicha, Moreno, & Kutas, 2004), and together imply that people's predictions include the syntactic properties of words. However, another study suggested that they predict aspects of their phonology as well. De Long,

Urbach, and Kutas (2005) measured event-related potentials while participants read contexts such as (14):

14a. The day was breezy so the boy went outside to fly a kite.

14b. The day was breezy so the boy went outside to fly an airplane.

Not surprisingly, the unpredictable *airplane* in (14b) led to a larger N400 effect than the predictable *kite*. The interesting finding was that this effect also occurred on the preceding article, indicating that readers were surprised to encounter *an* and therefore had predicted *kite* and, in particular, that it began with a consonant.

Another kind of evidence comes from situations in which listeners predict meanings when interpreting sentences about scenes, as measured by their eye movements to potential entities or events in the scenes. Altmann and Kamide (1999) found that participants tended to start looking at edible objects more than inedible objects when hearing *the man ate the* (but did not do so when *ate* was replaced by *moved*). This suggests that people predict entities by combining linguistic and nonlinguistic context (although it is possible that they are predicting meaning but not the grammar or the sound of the upcoming word). Further studies indicate that predictive eye movements depend on the meaning of the whole context, not just the meaning (or lexical associates) of the verb (Kamide, Altmann, & Haywood, 2003). Similarly, Knoeferle, Crocker, Scheepers, and Pickering (2005) had participants listen to *the princess washes apparently the pirate* (in German) while viewing a picture of a princess washing a pirate and a fencer painting the princess. They found that participants tended to look at the pirate before hearing *pirate*, thereby indicating that they predicted the event (i.e., the princess washing the pirate).

What system might be responsible for such predictions? One possibility is that there is a special-purpose system for making linguistic predictions during comprehension. But a simpler (and presumably less costly) alternative would be to use aspects of the production system to make predictions. When speakers produce language, they need to generate representations of the meanings, grammatical forms, and sounds that make up the planned utterance (e.g., Levelt, 1989). It is easy to see how parts of this system could be utilized during comprehension. While encountering *The tired mother gave her dirty child a...*, the reader could use the production system to construct representations of the meaning of *bath*, its grammatical category (noun), and its sounds. These representations can then be matched off against the analysis of the input, so that comprehension is facilitated if *bath* actually occurs. More specifically, Garrod and Pickering (2009) argued that comprehension involves imitating what has been heard using the production system, and then using those representations to make predictions. Informally, listeners ‘replay’ what they hear and then work out what they would say next.

So what is the evidence that such prediction does use production mechanisms? First, there is direct evidence for involvement of articulation in speech comprehension, with listeners activating appropriate muscles in the tongue and lips while listening to speech but not during nonspeech (Fadiga, Craighero, Buccino, & Rizzolati, 2002; Watkins, Strafella, & Paus, 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). Other fMRI studies have also demonstrated a large overlap between the cortical areas active during speech production and those active during passive listening (Pulvermüller et al., 2006; Wilson, Saygin, Sereno, & Iacoboni, 2004). These findings suggest that comprehension activates the production system and leads to covert imitation.

Recently, Federmeier (2007) proposed a similar account on the basis of evidence about the complementary roles of the hemispheres during comprehension. She argued that predictive processing is largely localized to the left hemisphere, while the right hemisphere concentrates on integrative processing. For example, Federmeier and Kutas (1999) presented implausible words that were semantically related to high-cloze completions of sentences in either the left or the right visual field. Event-related potential (ERP) data indicated a reduced effect of contextual congruency for the semantically related violations (e.g., *pin*es for *pal*ms) when compared with semantically unrelated violations (e.g., *tul*ips for *pal*ms), localized to the left hemisphere. They argued that this reflected semantic prediction processes associated with the left but not right hemisphere. Federmeier argued that language production is localized to the left hemisphere and hence suggested that language production circuits may be required for semantic prediction.

Another reason to suspect that comprehension uses the production system comes from the evidence of the pervasiveness of spontaneous (overt) imitation at many linguistic levels, which implies that people construct imitative plans at the relevant stages in the production process. Much of this evidence comes from studies of dialog and has already been noted (e.g., Branigan et al., 2000; Brennan & Clark, 1996; Garrod & Anderson, 1987), but there is also evidence for linguistic imitation in the absence of an interlocutor (e.g., Bock, Dell, Chang, & Onishi, 2007). For imitation to be used by the comprehension system, it must occur very quickly and result from automatic processes rather than conscious decisions. In fact, there is clear evidence that phonological or acoustic imitation is extremely rapid (Fowler, Brown, Sabadini, & Weihing, 2003) and the faster the imitation, the more faithful it is (Goldinger, 1998). Furthermore, interlocutors are almost entirely unaware that they imitate each other's grammar or choice of words (see Pickering & Garrod, 2004).

On the basis of such evidence, Garrod and Pickering (2009) proposed that listeners use the production system as part of an emulator (a forward model that operates in real time), which predicts the speaker's utterance (Grush, 2004). Although much of the evidence we have cited for production-based emulation comes from monolog, such emulation may be particularly important in dialog in two ways. On the one hand, the nature of dialog leads to particularly strong activation of the production system. On the other hand, emulation may be especially useful for achieving alignment. These two issues are of course related.

First, dialog involves regular switching between comprehension and production. When the speaker's turn finishes, the addressee is likely to take the floor. In some forms of dialog, speaker and addressee constantly change roles. In addition, the addressee may provide "backchannel" feedback during the speaker's turn, such as assertions (*yes, go on*, e.g., 7, 9, 11 above) or queries (*eh?*, *who?*). Such turn-taking and feedback does not occur in monolog (by definition). Thus, the production system is constantly activated during dialog, and hence it may be more active during comprehension of dialog than comprehension of monolog.

From a slightly different perspective, we can see that an addressee must be constantly prepared to respond. In some conversational settings, a response is normatively required (otherwise the addressee is flouting the rules of turn-taking; see Sacks et al., 1974). The simplest example of this is a question directed solely (and nonrhetorically) at the addressee, which requires some response. In monolog, the addressee knows that his only job is to comprehend what he is hearing. He may use production mechanisms to facilitate that comprehension, but he does not have to make overt responses.

Indeed, an addressee contrasts with other individuals who may be present in the dialog setting (e.g., Clark, 1996). Researchers distinguish different roles, typically speaker, addressee, side-participant, and nonparticipants (licensed overhearers and eavesdroppers). Importantly, the speaker addresses the addressee or addressees (e.g., by asking a question). A side-participant can respond but is not required to do so. Thus, we might expect the addressee to predictively process to a greater extent than side-participants or overhearers. In accord with this, Branigan, Pickering, McLean, and Cleland (2007) had speakers describe cards to addressees (as in Branigan et al., 2000) in the presence of a side-participant. They found that addressees were more likely to repeat the grammatical form used by the speaker than were side-participants. This suggests that addressees predicted the grammatical form used by the speaker in order to prepare a response to a greater degree than side-participants.

Additionally, Richardson, Dale, and Kirkham (2007a) compared gaze patterns in monolog and dialog. In monolog, addressees looked at pictures of the characters under discussion about 2 s after speakers; but in dialog, addressees and speakers looked at the pictures at about the same time. We propose that the addressees in dialog emulated their partners to a greater extent than addressees in monolog.

Second, emulation can assist with aspects of dialog that have no equivalent in monolog. In particular, comprehenders can use their emulation of the production system to assist in coordination (i.e., the smooth flow of the joint activity of dialog). Most dialog is ‘‘internally managed,’’ so that interlocutors themselves have to decide when it is appropriate to speak. They are remarkably good at this, because inter-turn intervals are extremely close to 0 ms in most exchanges. This means that the addressee knows precisely when it is appropriate to start speaking. To do this, the addressee can use the emulator to determine (in effect) how long an utterance the addressee would produce at that point, and assumes that the speaker will behave similarly. In accord with this, De Ruiter, Mitterer, and Enfield (2006) found that participants were very accurate at predicting the end of speakers’ utterances. Interestingly, their predictions remained accurate when the intonation was removed and the words were left but not vice versa. These findings follow from our account because the phonological, syntactic, and semantic predictions are largely driven by the words rather than by the intonational contour.

One potential problem is that interlocutors may speak at different rates, in which case the addressee’s emulations would not accurately predict the speaker’s actual productions. However, the interactive-alignment account predicts that interlocutors align their speech rate, and Giles, Coupland, and Coupland (1992) provide experimental evidence for such alignment. In a substantial corpus of both telephone and face-to-face conversations, ten Bosch, Oostdijk, and de Ruiter (2004) found that there was a very strong correlation between the

inter-turn interval between Interlocutor *A* and Interlocutor *B* and the inter-turn interval between Interlocutor *B* and Interlocutor *A*. In addition, Wilson and Wilson (2005) argued that interlocutors align their rate of syllable production and rapidly come into counterphase so that one interlocutor is ready to speak when the other is inhibited. This explains why an addressee rarely starts to speak at the same time as the original speaker, but why two addressees can start to speak at the same time. Emulation can bring this about and can thus be used to facilitate turn-taking.

Finally, the processor may emphasize emulation in dialog rather than monolog because dialog is in general more predictable than monolog. For example, it tends to be much more repetitive (see Pickering & Garrod, 2004). Clearly, the more repetitive a text, the more the advantage of emphasizing prediction over bottom-up analysis. Such repetitiveness occurs at many different linguistic levels, not just words. An efficient processor will therefore give relatively more weighting to the emulator versus the analysis of the input.

Although our emphasis throughout this discussion is on the use of production-based emulation by the addressee, it is also possible that the speaker uses such emulation to predict potential responses, either to questions or to other utterances that invite a response. For example, a parent might ask a child *Have you tidied your room?* and use the production system to predict the response *No*. The parent might even use this to prepare a further utterance (*Well, do so then*). In this case, the child's response is highly predictable; it does not matter that the parent's prediction derives from an utterance by the parent rather than an utterance of the child. This analysis is compatible with interpretations of contributions to the dialog that emphasize their anticipatory quality (e.g., Linell, 1998, Chapter 9).

6. Summary and conclusion

This paper explored how dialog mechanisms can be understood in terms of joint action involving at least two interlocutors. First, we argued that dialog is a joint action in which the participants' goal is to establish aligned representations of what they are talking about. Second, we argued that alignment occurs at many other levels of linguistic representation and comes about largely as the result of automatic processes based on imitation. We then argued that alignment also occurs for nonlinguistic aspects of communication. In this way, communication during dialog can be seen as a cline involving both linguistic and non-linguistic alignment. Finally, we argued that alignment comes about in part from prediction, both for comprehension and production. Indeed, communication during dialog represents a particularly rich form of joint activity.

Note

1. Notice that linguistic alignment can also enhance nonlinguistic alignment; for instance, remote cell phone users tend to synchronize their gait more during spontaneous conversations than when taking turns to describe pictures to each other

(Murray-Smith, Ramsay, Garrod, Jackson, & Musizza, 2007). This might, in turn, enhance linguistic alignment.

Acknowledgments

The writing of this paper was supported in part by a grant from the UK ESRC and MRC (RES-060-25-0010) to the first author and UK ESRC (RES-062-23-0376) to both authors.

References

- Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*, 247–264.
- Bavelas, J. B., Coates, L., & Johnson, T. (2000). Listeners as co-narrators. *Journal of Personality and Social Psychology*, *79*, 941–952.
- Bock, K., Dell, G. S., Chang, F., & Onishi, K. (2007). Structural persistence from language comprehension to language production. *Cognition*, *104*, 437–458.
- ten Bosch, L., Oostdijk, N., & de Ruiter, J. P. (2004). Durational aspects of turn-taking in spontaneous, face-to-face and telephone dialogues. In P. Sojka, I. Kopčec & K. Pala (Eds.), *Proceedings of the 7th International Conference, TSD 2004. Lecture Notes in Artificial Intelligence LNCS/LNAI 3206, Brno, Czech Republic, September 2004* (pp. 563–570). Berlin: Springer-Verlag.
- Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic coordination in dialogue. *Cognition*, *75*, B13–B25.
- Branigan, H. P., Pickering, M. J., McLean, J. F., & Cleland, A. A. (2007). Syntactic alignment and participant role in dialogue. *Cognition*, *104*, 163–197.
- Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 1482–1493.
- Brown-Schmidt, S., & Tanenhaus, M. K. (2008). Real-time investigation of referential domains in unscripted conversation: A targeted language game approach. *Cognitive Science*, *32*, 643–684.
- 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.
- Clark, H. H. (1996) *Using language*. Cambridge, England: Cambridge University Press.
- Clark, H. H., & Fox Tree, J. E. (2002). Using *uh* and *um* in spontaneous speaking. *Cognition*, *84*, 73–111.
- Cleland, S., & Pickering, M. J. (2003). The use of lexical and syntactic information in language production: Evidence from the priming of noun-phrase structure. *Journal of Memory and Language*, *49*, 214–230.
- De Long, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*, 1117–1121.
- De Ruiter, J. P., Mitterer, H., & Enfield, N. J. (2006). Projecting the end of a speaker's turn: A cognitive cornerstone of conversation. *Language*, *82*, 515–535.
- 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.
- Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, *44*, 419–505.
- Federmeier, K. D., & Kutas, M. (1999). Right words and left words: Electrophysiological evidence for hemispheric differences in meaning processing. *Cognitive Brain Research*, *8*, 373–392.
- 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.

- Fox Tree, J. E., & Clark, H. H. (1997). Pronouncing “the” as “thee” to signal problems in speaking. *Cognition*, 62, 151–167.
- Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, 27, 181–218.
- Garrod, S., & Pickering, M. J. (2004). Why is conversation so easy? *Trends in Cognitive Sciences*, 8, 8–11.
- Garrod, S., & Pickering, M. J. (2007). Alignment in dialogue. In G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 443–451). Oxford, England: Oxford University Press.
- Giles, H., Coupland, N., & Coupland, J. (1992). Accommodation theory: Communication, context and consequences. In H. Giles, J. Coupland & N. Coupland (Eds.), *Contexts of accommodation* (pp. 1–68). Cambridge, England: Cambridge University Press.
- Goffman, I. (1981). *Forms of talk*. Philadelphia: University of Philadelphia Press.
- Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105, 251–279.
- Grice, H. P. (1957). Meaning. *Philosophical Review*, 66, 377–388.
- Grush, R. (2004). The emulation theory of representation: Motor control, imagery, and perception. *Behavior and Brain Sciences*, 27, 377–435.
- Hartsuiker, R. J., Pickering, M. J., & Veltkamp, E. (2004). Is syntax separate or shared between languages?: Cross-linguistic syntactic priming in Spanish-English bilinguals. *Psychological Science*, 15, 409–414.
- Hatfield, E., Cacioppo, J. T., & Rapson, R. L. (1994). *Emotional contagion*. Cambridge, England: Cambridge University Press.
- 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, 133–156.
- 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.
- Kraut, R. E., Lewis, S. H., & Swezey, L. W. (1982). Listener responsiveness and the coordination of conversations. *Journal of Personality and Social Psychology*, 43, 718–731.
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Linell, P. (1998). *Approaching dialogue: Talk, interaction, and contexts in a dialogical perspective*. Amsterdam: Benjamins.
- Markman, A. B., & Makin, V. S. (1998). Referential communication and category acquisition. *Journal of Experimental Psychology: General*, 127, 331–354.
- Murray-Smith, R. D., Ramsay, A., Garrod, S., Jackson, M., & Musizza, B. (2007). Gait alignment in mobile phone conversations. In A. D. Cheok & L. Chittaro (Eds.), *Proceeding of MobileHCI 2007* (pp. 214–221). AMC International Conference Proceeding Series, Vol. 309, Singapore.
- Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27, 169–225.
- Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences*, 11, 105–110.
- Prinz, W. (1990). A common coding approach to perception and action. In O. Neumann & W. Prinz (Eds.), *Relationships between perception and action: Current approaches* (pp. 167–201). Berlin: Springer-Verlag.
- 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, 7865–7870.
- 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, 1045–1060.
- Richardson, D. C., Dale, R., & Kirkham, N. Z. (2007a). The art of conversation is coordination: Common ground and the coupling of eye movements during dialogue. *Psychological Science*, 18, 407–413.

- Richardson, M. J., Marsh, K. L., Isenhower, R. W., Goodman, J. R. L., & Schmidt, R. C. (2007b). Rocking together: Dynamics of intentional and unintentional interpersonal coordination. *Human Movement Science*, 26, 867–891.
- Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50, 696–735.
- Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences*, 10, 70–76.
- 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, 326–332.
- Traxler, M. J., & Gernsbacher, M. A. (2006). *Handbook of psycholinguistics* (2nd ed.). San Diego, California: Academic Press.
- 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, 443–467.
- Watkins, K., & Paus, T. (2004). Modulation of motor excitability during speech perception: The role of Broca's area. *Journal of Cognitive Neuroscience*, 16, 978–987.
- Watkins, K., Strafella, A. P., & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia*, 41, 989–994.
- Wicha, N. Y. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience*, 16, 1272–1288.
- 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, 701–702.
- Wilson, M., & Wilson, T. P. (2005). An oscillator model of the timing of turn-taking. *Psychonomic Bulletin & Review*, 12, 957–968.