Authors’ Response

The interactive-alignment model: Developments and refinements

Martin J. Pickeringa and Simon Garrodb

aDepartment of Psychology, University of Edinburgh, Edinburgh EH8 9JZ, United Kingdom; bDepartment of Psychology, University of Glasgow, Glasgow G12 8QT, United Kingdom. Martin.Pickering@ed.ac.uk simon@psy.gla.ac.uk

Abstract: The interactive-alignment model of dialogue provides an account of dialogue at the level of explanation normally associated with cognitive psychology. We develop our claim that interlocutors align their mental models via priming at many levels of linguistic representation, explicate our notion of automaticity, defend the minimal role of “other modeling,” and discuss the relationship between monologue and dialogue. The account can be applied to social and developmental psychology, and would benefit from computational modeling.

The target article set out to show how it would be possible to develop a theory of interactive language processing at the level of explanation normally associated with cognitive psychology. In our theory, successful communication involves the alignment of interlocutors’ representations. We proposed that each level of representation becomes aligned via an automatic process that we treat as a form of priming, and that alignment at one level automatically strengthens alignment at other levels. The role of conscious or deliberate strategies involving explicit reasoning about the mental states of one’s interlocutor is comparatively small in our account.

Our commentators have raised a number of insightful points that have caused us to refine our proposals. Many commentators have focused on the nature of the alignment process. At a basic level, they consider whether alignment is the primary mechanism leading to conversational success, to what extent it is automatic, and whether it can be explained by a single mechanism at all levels and in all contexts. Commentators have also questioned our downplaying of “other modeling” in ordinary conversation and our claims about the nature of the difference between monologue and dialogue. In responding to these and other comments, we have divided our reply into eleven sections whose order roughly follows that of the topics raised in our target article.

R1. To what extent do interlocutors align?

Perhaps the most basic issue about our model is whether interlocutors actually align their situation models, or, less dramatically, whether they align to the extent that we claim they do. Schober proposes that interlocutors may be much less aligned than they appear even when they believe that they have understood each other. Of course, this would not matter if it solely concerned rare cases of genuine misunderstanding (e.g., when two interlocutors refer to different people called John); see also Branigan, who points out that communication may be “successful” in some sense even when there is some misunderstanding. But Schober argues that misalignment is endemic to dialogue. His comments relate particularly to the interpretation of referring expressions with respect to the discourse model. He draws on examples from surveys where respondents interpret terms in ways that are very different from those intended by the survey composers. Our response is that such surveys do not constitute dialogue: The composers construct the survey, and the respondents then respond. There is no feedback, no possibility for repair, and hence no interactive alignment. Schober also raises the important point that people need not necessarily fully interpret expressions (Clark & Wilkes-Gibbs 1986). In fact, full interpretation probably does not always occur in the comprehension of monologue (Barton & Sanford 1993; Frazier & Rayner 1990; Frisson & Pickering 1999; 2001; cf. Sanford & Sturt 2002), with people often not determining the precise sense of referring expressions (e.g., newspaper meaning an object vs. a day’s edition), and there is no reason to assume that dialogue is any different. We suspect that both producers and comprehenders determine meaning to the extent necessary for current purposes, and that one way in which interlocutors align is by each processing referring expressions to equal depth.

R2. What precisely are they aligning?

Several commentators appear concerned with the question of what exactly is being aligned within our model. At the “lower levels” of phonology, syntax, the lexicon, and so on, interlocutors presumably align the representational content of each of those levels (phonemes, syntactic structures, lexical items, etc.), but it is perhaps less clear what they align at the level of the situation model. In the target article, our intention was to argue for alignment of structural aspects of the situation model, as exemplified by our example of reference frames. Some of our commentators assume that we are referring to the content of the situation model. The questions about alignment of content are much more difficult, and we shall try to explain the issues below.

In our account, interlocutors align on representations relevant to the dialogue. These include lexical, semantic, and syntactic representations, but also the situation model. So if, at a given point in a conversation, one interlocutor has a situation model containing two individuals, Mary and John, with Mary in focus, with each at different locations, and so on, then the conversation will be successful to the extent that the other interlocutor constructs the same situation model. Of course, one interlocutor can now introduce another character (or a new relation between the existing characters) – indeed, introducing new information is central to any conversation that is not entirely repetitive. To do this, the speaker draws upon his knowledge (typically using long-term memory) and adds information to his situation model. The effect of the alignment is that the listener updates his model so that it remains similar to that of the speaker. For example, the listener will interpret ambiguous words and utterances in the way that the speaker has employed them.

A much bolder claim is that the choice of new topics is affected by alignment. We did not make this claim in the target article, although we believe that it is true to some extent. For example, if one interlocutor refers to the couch, then the other is more likely to refer to the couch as well (Breen & Clark 1996; Garrod & Anderson 1987). As a result of this, the use of couch presumably activates knowledge about couches, and hence makes it more likely that the interlocutor will talk about couches rather than some
other topic. To this extent, alignment is surely unsurprising (and simply amounts to the claim that interlocutors will persist with particular topics).

It may also be that interlocutors align on particular styles of reasoning or accessing of knowledge. For instance, if one interlocutor is engaged in a careful search of long-term memory, then the other will tend to behave similarly (e.g., if you play a general-knowledge game seriously, then I am likely to do so too). Alignment on style of reasoning is relevant to the construction of the situation model (cf. Gentner & Markman 1997), but it takes us beyond the scope of the target article, just as nonlinguistic imitation more generally does (e.g., Chartrand & Bargh 1999). For now, our goals are limited to understanding linguistic factors that assist in the alignment of situation models.

Cutting questions our characterization of interactive alignment as reflecting a direct link between interlocutors. He suggests that it only has an indirect effect on the language processes themselves. To answer this comment we need to clarify how interactive alignment relates to language processing. As we have said, our contention is that interactive alignment (and in particular the automatic alignment channels) affects the structures used in production and interpretation rather than directly determining the content of production and interpretation. In other words, we assume that alignment provides an explanation of the manner in which interlocutors produce and interpret contributions. So we propose that alignment channels only produce a direct link between the structures that the interlocutors use in language processing. Hence, the alignment process is automatic and direct, even though it does not determine exactly what the speaker produces (as this depends on his long-term knowledge) or how the addressee interprets what he hears “beyond” the level of the situation model.

Other commentators also query whether we have specified the appropriate mechanism for alignment of situation models. They raise this concern in relation to two more specific topics: whether there is one alignment mechanism or several (Branigan, Glucksberg, and Markman, Kim, Larkey, Narvaez, & Stilwell [Markman et al.]), and whether (or in what sense) alignment is automatic (Krauss & Pardo). All of these commentaries, in some sense, are concerned with the issue of how alignment could affect the content of situation models. We address these in the following two sections.

R3. The mechanisms of alignment

Several commentators question the details of the interactive alignment mechanism itself and point out that we have not fully specified a mechanistic account. Brown-Schmidt & Tanenhaus make a general plea for modeling, which we fully agree with (see sect. R11). Goldinger & Azuma argue that we do not give a detailed characterization of the process by which alignment comes about. We have no commitment to interactive-activation models and are open to the suggestion that Grossberg’s (1980) adaptive-resonance theory may provide an appropriate framework for the interactive-alignment account.3

Beyond this, two somewhat different issues are raised. Some commentators argue that we assume alignment is based on transient activation, and they propose instead that it is based on facilitated memory retrieval or implicit learning. Others claim that we are wrong to assume a unified account for all levels of alignment.

Kaschak & Glenberg argue that alignment is not due to priming but rather to a facilitated memory retrieval mechanism. In response, we note that the interactive-alignment model is specified at a functional level and makes no commitment to specific mechanisms, and that we use the term “priming” to refer to both transient activation and facilitation in memory-based accounts. Our model attempts to capture the way in which representations used for both production and comprehension automatically become aligned as a consequence of the process of interaction. These representations may be subject to transient activation or, instead, there may be enhancement of the mechanisms underlying their retrieval from memory (as envisaged by Kaschak & Glenberg).

Perhaps more likely, there may be two separate mechanisms involved in alignment. For example, some recent accounts of syntactic priming are based on implicit learning (Bock & Griffin 2000; Chang et al. 2000), whereas some are based on activation of grammatical nodes (Hartsuiker et al. 2004; Pickering & Branigan 1998). Some experimental research finds clear evidence for long-term priming that is largely unaffected by intervening material (Bock & Griffin 2000; Hartsuiker & Westenberg 2000), whereas others show rapid decay (Branigan et al. 1999; Levelt & Kelter 1982; Wheelton & Smith 2003). Most likely, different tasks and sentence types lead to very different time-courses of priming. Although most of this work does not involve dialogue (except Levelt & Kelter 1982), under our account we would expect similar patterns of results to occur in dialogue. We therefore suggest that transient activation explains some aspects of alignment, and memory-based mechanisms explain other aspects of alignment. In section R9 below, we suggest that alignment due to routinization is likely to involve the establishment of memory traces for semi-fixed expressions.

Schiller & de Ruiter argue that interactive alignment involves storing and re-using selected fragments from previous utterances (see sect. R9); this constitutes a specific version of a memory-based account. However, their argument is based on the claim that priming is insufficient to account for interactive alignment because syntactic priming effects are too weak. In fact, the 10–20% effects that they refer to, occur in monologue. In dialogue, our studies have shown 55% priming effects when the verb is repeated (Branigan et al. 2000) and up to 47% with a rare structure when the noun is repeated (Cleland & Pickering 2003). Likewise, lexical entrainment almost always occurs for ambiguous words (Brennan & Clark 1996; Garrod & Anderson 1987). In our model, percolation effects between levels also increase the degree of alignment, and extended dialogue iteratively reinforces alignment.

A number of commentators question whether alignment operates in the same way at all levels in our model. Markman et al. argue that there are different requirements on alignment at the different levels. In particular, they separate the situation model from lower levels of linguistic representation. We agree that the structural alignment process they identify may well be appropriate at the level of the situation model, because models reflect complex higher order relations between elements (see sect 2.2 of the target article). However, we disagree with their argument that, unlike
lower level representations, situation models have to be partially misaligned either to ensure that given-new conventions are followed, or for the maintenance of common ground. We propose that these requirements can be fulfilled through the implicit common ground which does not differentiate between the speaker’s and listener’s situation models.

Branigan also separates the situation model from other levels, but for reasons that differ from those of Markman et al. In our terms, she accepts channels of alignment at syntactic, lexical, and morpho-phonological levels but not at the level of the situation model, because she believes that utterances do not provide direct evidence about the situation model. She claims that I am in row two provides direct evidence about lower levels, whereas the listener has to interpret the utterance (presumably, by using background knowledge) in order to construct the situation model. We disagree with this, because all levels of analysis require a combination of top-down and bottom-up information. For example, resolving phonemes, ambiguous words, or syntactically ambiguous utterances requires the use of context. It is therefore wrong to assume that only the level of the situation model is “abstract.” We therefore see no reason to assume that channels of alignment are used only at lower levels, nor do we see any reason to alter our assumption that alignment at lower levels leads to alignment at the level of the situation model.

Warren & Rayner argue that the priming link between individuals’ situation models must be different from that for lower levels. This is because interlocutors do not necessarily begin dialogues with similar situation representations and so alignment has to be built up over a period of interaction. Again, we see no fundamental difference between situation models and lower levels in this respect – alignment at all levels is built up, though the rate of alignment may differ at different levels. Additionally, Warren & Rayner question how conflicts in alignment at different levels are resolved (e.g., when aligning on the same name, “John” might produce a semantic misalignment in contexts where there are two Johns present). In fact, the issue was briefly discussed in the target article where we argued that alignment at the level of the situation model would override alignment at lower levels (target article, endnote 4). Adopting a particular situation model will influence the way a speaker frames almost everything he says, whereas adopting a particular word or syntactic structure will only affect the subsequent choice of that word in preference to another or influence the subsequent use of that particular syntactic structure. Because the situation model is so pervasive, it will be constantly reinforced in implicit common ground, and misalignment at this level will be more likely to trigger interactive repair. This suggests that the time-course of priming at the level of the situation model may be long-lasting, whereas priming at low levels, such as phonology, may be much more short-lived. Priming at the syntactic level might be intermediate in duration, or depend more on its precise context (as suggested above). It therefore might be the case that priming of the situation model depends primarily on memory representations, whereas priming at low levels might depend primarily on transient activation. All of this, however, requires detailed modeling.

We also believe that routinization plays an important role in reinforcing the links between lower and higher levels of representation. We take this up in section R9.

R4. What it means for interactive alignment to be an automatic process

The commentators raise two important issues about automaticity that require discussion. Krauss & Pardo argue against the idea that alignment can be accounted for in terms of automatic priming between interlocutors. Shintel & Nusbaum argue that speech comprehension processes may be far from automatic in dialogue. To answer these concerns we need to first explicate our notion of automaticity, and, second, indicate what we assume to be automatic.

Our notion of automaticity is derived from the perspective of perception-action relationships (e.g., Hommel et al., 2001) and, more particularly, social cognition and social cognitive neuroscience (e.g., Dijksterhuis & Bargh, 2001, Harley & Clater, in press). Just as Dijksterhuis and Bargh argue for an automatic perception-behavior expressway, we propose that the alignment channels are automatic (see sect. 3.2) – they operate without any intermediary decision process. Hence, the alignment process is automatic. To be more explicit, we propose that the automaticity of alignment may take place at what Bargh (1989) calls the post-conscious level, whereby automaticity requires awareness of the stimulus when it originally occurred. This means that interlocutors have to attend to what the other is saying in order for automatic alignment to occur. Dijksterhuis and Bargh (2001, p. 29) also argue that automatic social influences can be inhibited when they conflict with current goals and purposes. We suggest that the same is true for interactive alignment (see Garrod & Pickering, 2004). For example, if a maze game player wants to try a new description scheme because he has failed to understand the last description from his interlocutor (see sect. 2.1 of the target article), then this high level goal of introducing a new scheme may inhibit low level alignment arising from what his interlocutor has just said. However, in a similar vein to Dijksterhuis and Bargh, we predict that overriding alignment is going to be more difficult (or effortful) than adopting alignment. Additionally, this postconscious notion of automaticity can explain why alignment is affected by partner-specific factors (e.g., Branigan et al., submitted; Metzing & Brennan, 2003), without invoking additional mechanisms such as “other modeling.” It is also presumably relevant to many of the factors that affect the extent of speech accommodation (Giles et al., 1992). In general, we expect that rate of alignment may be affected by social factors even when the interlocutors are unaware that they are aligning. There is evidence for such alignment outside language (Eppey & Gilovich, 1999; Lakin & Chartrand, 2003), and we expect it also to occur in language.

Krauss & Pardo agree with our claim that communication entails the alignment of situation models, but express that it does not principally take place via automatic priming. For example, they point to evidence that speakers accommodate to their listeners. This presents no problem according to the above conception of automaticity, which allows inhibition or facilitation by social factors. Glucksberg raises an interesting case, involving a difficult dialogue with a non-cooperative teenage son, in which degree of alignment may be reduced.

Shintel & Nusbaum argue that speech comprehension processes may be far from automatic in dialogue. We are quite happy to accept this general point but see no prob-
lems for our proposal. In our account, the process of aligning the structures used in comprehension (and production) is automatic, but other aspects of comprehension (and production) are presumably not automatic. Additionally, their conception of automaticity is that it “implies a passive process in which the input is processed in an invariant, inflexible manner, regardless of the beliefs and expectations of the listener.” This is not the notion of automaticity that we intend, and we hope that the above discussion of Bargh (1989) helps to make this clear.

Our conception of automaticity also differs from a Skinnerian one, as suggested by Pear. Crucially, we assume that alignment is not due to reinforcement, just as Dijksterhuis and Bargh (2001) assume for the perception-behavior expressway. Instead, alignment follows from a primitive tendency to imitate that does not appear to be learned (e.g., Metzoff & Decety 2003). However, our account does share certain features with Skinner’s (1957) account, in particular that alignment implicates low-level learning mechanisms.

R5. Parity

One concern is whether there is true representational parity between production and comprehension. Ferreira describes experiments in which participants plan to produce utterances that they know to be ungrammatical (i.e., participants do not simply make errors). She assumes that interlocutors use and understand such utterances during dialogue (which is almost certainly correct) and suggests that comprehenders in dialogue would regard them as illicit. Whereas it is possible that there are differences between monologue and dialogue with respect to judgments of grammaticality, we accept that such differences are unlikely. In her experiments, we suggest that speakers realize they are producing something ungrammatical, but do so anyway because they cannot think of any other way of saying what they want to say. As long as this realization takes place within the production system (i.e., does not purely occur during self-monitoring), there does not appear to be a problem for the parity assumption. Compare sports commentators who sometimes cannot identify a player at the point when they need to produce the utterance (“About to kick the ball, Smith”), which listeners might well regard as illicit. This account seems more likely than a real disconnect between grammars in comprehension and production. However, if there is a disconnect for some highly specific constructions, it merely leads to a very slight weakening of the parity assumption, not its abandonment.

Ginzburg argues that the interpretation of the same sequence of words can change according to whether it represents a single contribution from one speaker or two contributions from different speakers:

1. A: Which members of our team own a parakeet? A: Why? (= Why own a parakeet?)
2. A: Which members of our team own a parakeet? B: Why? (= Why are you asking which members of our team own a parakeet?)

He suggests that our interactive alignment mechanism cannot account for the fact that Why? has a different interpretation in interactions (1) and (2). This is an interesting observation, but the difference in interpretation between (1) and (2) hinges on the dialogue move (e.g., questioning, answering, checking, informing) being performed at that point. Because dialogue moves are generally associated with particular speakers, it is obviously crucial that interlocutors monitor the source of an utterance when interpreting it (as also follows from results like those of Metzing & Brennan 2003). For example, the speaker treats a question from his interlocutor differently from the way he would treat a question from himself. We accept that interlocutors can monitor the source of a contribution (i.e., they can differentiate between what they are saying and what their partner is saying) and can take this into account in their interpretation at the level of the dialogue move.

Cutting proposes parity for semantic but not phonological representations on the basis of picture-word interference experiments. From his brief description, we suggest that participants process the words that they actually produce both semantically and phonologically, but that they process the words that they are told to ignore semantically but not phonologically (or at least not to a sufficient depth to affect priming). Krauss & Pardo also question evidence for phonological alignment (and by implication for phonological parity). Although we accept that Goldinger (1998) does not directly demonstrate phonological alignment, recent evidence does support parity between production and comprehension at this level (Fowler et al. 2003).

Kempen defends a more radical proposal that parity comes from the symmetry between production and parsing processes. In her Dynamic Syntax account of parsing, syntactic information is combined with lexical information, which define semantic interpretations that are built up word-by-word. Production is assumed to work in essentially the same way. Hence, she sees interactive alignment as operating at the level of the production and parsing processes themselves. This is a challenging linguistic proposal, but it would need explicit modeling before it could be incorporated into a mechanistic account of language processing in dialogue.

R6. Is it only misunderstanding that drives interactive repair?

One concern is whether interactive repair is driven primarily by comprehension failure, as we proposed in section 4.3 of the target article. Healey points out that even in the context of Garrod and Anderson’s (1987) maze-game dialogues, interlocutors change their description scheme in a systematic fashion (e.g., shifting from a path or figural scheme to a line or matrix scheme). He argues that it is unlikely that this systematic shift can be accounted for only in terms of an interactive repair mechanism based on comprehension failure. Of course we recognize (see sect. 4.4 of the target article) that alignment does not depend only upon this process. There are many things that determine what people choose to say and even how they do so which go beyond the simple automatic mechanisms discussed in the target article. For example, the shift in description scheme that Healey mentions probably reflects two opposing pressures. Whereas the abstract line and matrix descriptions are more efficient over a period of time than figural or path descriptions (e.g., a line or matrix description involves fewer words and is not influenced by whether the position is near a salient point in the maze or lies in a salient pattern), they are more difficult to align (e.g., matrix descriptions depend upon alignment of the origin and of the counting conven-
tions used). So they can often not be used securely until there is a richer implicit common ground (e.g., repeated use of path descriptions which begin at one corner of the maze can lead to this corner being adopted as the origin for matrix description). We suspect that once the implicit common ground has become sufficiently rich to support the more abstract description, a speaker is more likely to adopt that scheme when he encounters a position that is particularly awkward to describe even when it requires a violation of alignment.

Note that this shift occurs without the speaker having to take account of the listener's knowledge. Healey therefore brings up an important general point, that interlocutors can go beyond interactive alignment and repair in ways that do not require other-modeling or the establishment of a full common ground. For example, a speaker can decide that a representational scheme is unnecessarily complex or a referring expression is unnecessarily long even if the interlocutors have aligned on that scheme or expression. Similarly, in preparing lectures, I might change how I am speaking on the basis of my knowledge of the audience (particularly, in preparing lectures, I might change how I am speaking on the basis of my knowledge of the audience (full larly, in preparing lectures, I might change how I am speaking on the basis of my knowledge of the audience (full common ground), but I might also do it on the basis that "Hang on, I'm not doing this efficiently, given my own resources – I am trying to remember too much and can't manage it." This might be argued to involve access to a second model of one's own mental state, which is therefore costly, but less costly than keeping track of full common ground. Such decisions require there to be some inhibition of the basic alignment process in light of a conflicting goal (see sect. R4). In conclusion, Healey's point reflects something that is additional to our account rather than in conflict with it (cf. Krauss & Pardo, who point out that not only mis-understanding drives accommodation).

R7. Other modeling

Although interlocutors undeniably do pay attention to each others' mental states on occasion, our contention is that such other-modeling is resource-intensive, essentially because it involves storing two representations: a representation of one's own state of knowledge, and a separate representation of one's partner's state of knowledge. We therefore believe that most of the process of alignment occurs via the interactive-alignment mechanism where other-modeling is not required. But we stress that other-modeling is not purely used for "difficult" cases of interactive repair when automatic processes fail to work. When a boy decides to tell his mother what happened at school today, he presumably realizes that his mother does not know about the event in question, and therefore that he knows something that his mother does not know. This explicit modeling of the difference between knowledge states leads to him running to tell his mother about the event, and does not follow from the failure of interactive alignment and interactive repair. Similarly, a bilingual decides which language to speak on the basis of his assumptions about which language his listener knows. However, the undeniable use of such "broad-brush" other-modeling does not mean that other-modeling is employed in a fine-grained way to explain detailed decisions about one's individual contributions to an ongoing dialogue.

In this context, Krauss & Pardo point to evidence that speakers sometimes modulate their utterances to take into account their knowledge of the listener: They produce more informative contributions when they perceive their addressees to be less knowledgeable about the relevant topic (see also Isaacs & Clark 1987). The evidence from Kingsbury (1968) shows that speakers do not simply pay attention to what they believe about their specific interlocutor but make inferences about how much such a person is likely to know on the basis of the evidence at hand, which is, in this case, made quite deliberately apparent to the speaker (e.g., the questioner frames a question to stress his ignorance of the city). In general, we suspect that speakers make a one-off decision based on such issues as the perceived expertise of their addressees about how to frame their contributions (e.g., the decision not to make any assumptions about local geographic knowledge). A teacher can be much less explicit in the common room than in the classroom, and a mother does not speak motherese to her friends. Such decisions need not remain fixed for the whole conversation (e.g., they might change when the speaker guesses that his addressee is not a local but then realizes he is mistaken). But such a change is very different from a continuous, dynamic process of utterance accommodation based on full common-ground inference, which we argue to be implausible for reasons of resource limitations (sect. 4.1 of the target article). We are therefore grateful to Krauss & Pardo for stressing that explicit modeling does not only occur when automatic processes fail to produce alignment, but we see no concern for our assumption that automatic mechanisms underlie alignment.

Fussell & Kraut argue that speakers with different views of a spatial scene take into account the listener's perspective, in effect modeling the listener's mental state. They describe a collaborative bicycle repair situation in which an expert helper guides a novice repairer. They note that when the repairer knows that he can be seen even when he cannot see his remote helper, he will use deictic expressions to describe the things in front of him (e.g., *See this piece*, while pointing at a cycle component), whereas the remote helper will not (e.g., *See the derailleur*). They argue that this is inconsistent with alignment and provides further evidence of other modeling. We are not convinced. We suspect that speakers in this situation prefer to use deictic expressions because they are shorter, do not require word finding, and so on. But deixis is not an option for the remote helper because he cannot point to anything. Instead, he has to fall back on more complex nondeictic descriptions. (One remote helper is quoted as saying in frustration, "If I could point to it, it's right there"; Kraut et al. 2003, p. 36.) So the circumstances may force the speaker to use a more complex nonaligned utterance. It is of course reasonable that alignment is broken under such circumstances, because it simply would not work. One important point this raises is that the tendency toward alignment is likely to be stronger under conditions where two interlocutors are placed in comparable environments. Presumably this reflects nonlinguistic contributions to linguistic alignment (see also the discussion of Dominy in sect. R1).

Nevertheless, we certainly agree with the general point that when communicators share a physical situation they take situational awareness into account in formulating utterances. But is this evidence for listener modeling? In the "side-by-side" situation described by Kraut et al. (2003, communicators use direction of gaze to establish joint attention, but the effect of one partner's point of gaze on the
other partner's focus of attention reflects low-level mechanisms which do not depend on inferences about the partner's mental state (see Schuller & Rossion 2001). And, because in this situation what is accessible to the speaker will usually be equally accessible to the partner (see sect. 4.1), an essentially egocentric approach will generally support successful communication without requiring speakers to model their listeners.

Schöber suggests that current evidence cannot distinguish two possibilities: that interlocutors only model each other's mental states under exceptional circumstances, and that interlocutors normally model each other's mental states and only fail to do so when under great cognitive load or when circumstances weigh heavily against doing so. We accept that current evidence does not distinguish between these two positions. However, our account assumes the use of fewer resources and is parsimonious (obviously, an account containing two mental models is harder to falsify than an account limited to one, just as a parallel account is harder to falsify than a serial account). Moreover, Schöber's proposal cannot hold for multiparty dialogues containing more than a small number of people, because it must become impossible to retain and regularly update a different mental model for each person. In general, cognitive psychology teaches us that constructing mental models is hard and holding onto different models at the same time is especially hard (e.g., Johnson-Laird 1983). We suggest that the paradox of how one can know when to model one's partner is easy to accommodate: Contributions to the dialogue will make it clear that alignment is breaking down, and if interactive repair does not solve the problem, the interlocutor is forced to assume that what his partner knows is likely to be different from what he knows. Even in such cases, it may be that interlocutors only model those differences between themselves and their partners that need to be modeled in order to allow the recovery of alignment.

Brennan & Metzing also criticize our assumption that interlocutors do not routinely employ full common ground. A fast-growing body of literature suggests that interlocutors sometimes do pay attention to each other's knowledge in comprehension and production (e.g., Hanna et al. 2003; Lockridge & Brennan 2000; Nadig & Sedivy, 2002) and sometimes do not (e.g., Brown & Dell 1987; Ferreira & Dell 2000; Keysar et al. 2003). It is too early to say precisely when such knowledge can affect processes of production and comprehension, but current evidence suggests both that interlocutors can immediately draw upon knowledge about differences between their own knowledge and their beliefs about their partner's knowledge, and that interlocutors can make egocentric decisions about production and comprehension. Most of these studies involve a fairly artificial situation in which the experimental subject is informed that his interlocutor may have knowledge about the situation that differs from his in quite specific ways. Experiments like that of Hanna et al. (2003) show that it can be straightforward to add one fact about your interlocutor — namely, that he does not have access to a particular piece of information that you have. Even in such cases, some egocentric behavior remains, as Hanna et al. acknowledge and Keysar et al. (2003) demonstrate. But adding one fact about your interlocutor's knowledge is quite different from maintaining a full representation of the interlocutor's situation model, and performing reasoning based on that model. Available resources do not normally allow interlocutors to constantly update models of each others' mental states. However, this does not lead to communicative breakdown because aligned interlocutors develop the same situation models.

In response to Brennan & Metzing, we stress that it was not our intention to commit to a two-stage account (e.g., Horton & Keysar 1996), in which other modeling occurs during revision but not during initial processing (whether production or comprehension). We note that Krauss & Pardo and Brown-Schmidt & Tanenhaus also interpret us as making this proposal, and accept that we did not make this very clear. Rather, we claim that "performing inferences about common ground is an optional strategy that interlocutors employ only when resources allow" (target article, sect. 4.2, para. 4).

We do not regard Metzing and Brennan's (2003) demonstration of partner-specific effects as problematic, and assume they can be explained in similar ways to Branigan et al.'s (2003) demonstration that syntactic alignment is sensitive to participant status (see sect. 2.3 of the target article). As we have pointed out in section R4 of this response, we assume that alignment is automatic at a postconscious level (Bargh 1989) and, hence, can be affected by a range of social factors from stereotype activation to participant status. A particular speaker is associated with a particular form, and breaking that association causes disruption. There is no need for other modeling to occur in this process of partner-specific lexical entrainment. The term conceptual pact appears to suggest that other modeling is used in lexical entrainment. If so, we would question whether it is generally appropriate.

R8. Routines

Schiller & de Ruiter propose that interactive alignment necessarily involves selecting stored fragments from previous utterances. This corresponds to our notion of routinization (see sect. 5 of the target article). We suspect that routinization comes about as a result of a longer lasting alignment mechanism based on memory retrieval rather than transient activation. This is because routines reflect multiple links between different levels of representation (e.g., they fix the relation between a word and its meaning, its syntactic form, and even its interpretation within a situation model) and it is difficult to imagine how this could be captured and routinized through purely transient activation. Rather than assume that routinization is the sole explanation of alignment, we suggest that it is a consequence of implicit learning but that transient activation also promotes alignment (see sect. R3). It may of course be that routines emerge from a resonance process, as Goldinger & Azuma suggest. In addition, because routinization works by linking levels of representation, it may explain how alignment percolates up from lower to higher levels (cf. issues raised by Warren & Rayner and Branigan, as discussed in sect. R3).

Within the interactive-alignment account, we regard routines as an extreme case of alignment, involving a fixed form and interpretation. It may be best to think of routinization as falling on a continuum, with expressions that contain some fixed elements (as in many of Kuiper's 1996 examples) being more or less "semi-routinized." Assuming that it is correct to regard alignment as a mixture of tran-
sient activation and implicit learning, we propose that the more routinized an expression is, the more it is best explained in terms of implicit learning – for the purposes of the conversation at least, the expression and its interpretation are stored and retrieved. Of course, if an expression becomes sufficiently entrenched, it may survive that conversation. Although other frameworks are no doubt possible, we regard Jackendoff’s (2002) account of fixed and semi-fixed expressions as an appropriate representational scheme for semi- and completely routinized expressions (see Pickering & Garrod, in press, for discussion).

R9. Self-monitoring

Schiller & de Ruiter question our claim that self-monitoring can occur at any level of linguistic representation that can be aligned. We did not claim that there is conclusive evidence for this hypothesis and we believe that careful empirical work is needed to distinguish our proposal from the proposal that monitoring works externally on sound and internally on phonological representations alone. However, we would query whether the reported evidence provides strong support for this alternative proposal. First, the comparative slowness of selecting a gender label in comparison to selecting the indefinite article in French may have many explanations, perhaps most likely that selecting between genders is a more abstract and difficult task than selecting between (very common) words. Second, the strong evidence for monitoring of various aspects of phonological representations is completely compatible with monitoring of other linguistic representations. Although some or all gender-congruency effects in picture-word interference tasks may really be determiner congruency effects (Schiller & Caramazza 2003), there is also considerable evidence that grammatical gender can be accessed when phonological form is not available (Badecker et al. 1995; Vigliocco et al. 1997). Therefore, it is at least plausible that people can directly monitor for errors of grammatical gender and indeed for other aspects of syntactic representations. If an utterance is ill-formed at different levels of representation simultaneously, we suspect that there may be a race between monitoring processes at these different levels, in which case it might not always be possible to detect monitoring that takes place at the “slower” level.

R10. On the difference between dialogue and monologue

A number of commentators argue that language processing in dialogue is not fundamentally different from that in monologue. For example, both Barr & Keysar and Glucksberg point out that the same basic language processes operate in monologue and dialogue so there is no principled difference between the two. We agree in the sense that the actual production and comprehension mechanisms are the same (at what we might term a “microlevel”). However, the radically different contexts in which they operate lead to very different results. For example, a speaker’s utterances are dramatically affected by the presence of the interlocutor – the speaker aligns with the interlocutor’s utterances via the mechanisms we have described. In this respect we argue that the language processing system is designed for dialogue rather than monologue. As a result, speakers have to learn special strategies to deal with monologue which are not required during dialogue processing.

We agree with Glucksberg that dialogue is not necessarily easier than monologue, and accept that contextual effects can be very strong in monologue. We propose that the priming mechanisms are ideally suited for dialogue. Presumably they have developed from imitation (Arbib, in press) and it may be that the organization of dialogue (e.g., time between turns) is optimal for the mechanisms of priming. Therefore, dialogue does not need to rely on nonautomatic inference. In contrast, monologue cannot use priming between interlocutors (by definition) and therefore has to rely on inference, other-modeling, and so on. Priming is of course present in monologue, but we contend that it is far less useful than in dialogue (e.g., repetition is much rarer in monologue than in dialogue; see sect. 5.1 of the target article). So we concur that there is not a principled distinction between dialogue and monologue, but at the same time we believe that the difference is less than is often suggested for dialogue than in monologue; see sect. 5.1 of the target article). So we concur that there is not a principled distinction between dialogue and monologue, but at the same time we believe that the difference is less than is often suggested.

Barr & Keysar appear to disagree with us more than we think they actually do. They are mistaken in assuming that we propose a categorical distinction between monologue and dialogue. In section 8 of the target article, we refer to a “dialogic continuum” with monologue at one end, and fully interactive dialogue at the other. We assume that the same mechanisms are present in dialogue and monologue (i.e., people do not set some processing “switch”). In true monologue, the speaker has no interlocutor to align with. He can of course align with himself and certainly does so (e.g., re-using the same word with the same meaning). We completely agree that dialogues go through various stages, with some involving rapid turn-taking (e.g., question answering) and some involving much more limited feedback (e.g., during narratives). Boden (1994) distinguishes between conversational phases and presentational phases in group discussion. These presentational phases are not monologues, as even minimal feedback affects them considerably (Bavelas et al. 2000).

Hence, we stress that monologue and dialogue lie on a continuum, and we predict that the degree of alignment will be affected by the position on the continuum. One important area for research is to consider the effects of dialogue genre on alignment (in which context we can regard monologue as particular genre). For example, Schegloff points to the importance of different speech-exchange systems (conversation, interview, giving a speech, etc.) in affecting the characteristics of the dialogue (e.g., turn-taking behavior, routinization). We predict that the rate and characteristics of alignment are not constant for all forms of dialogue, but will depend on the speech-exchange system. For example, forms of interaction that do not allow unconstrained feedback and where turn-taking is externally managed (e.g., interviews) will fail to employ the interactive repair mechanism to the extent that is possible in casual conversation.

R11. Extensions and discussion

Schegloff argues that our mechanistic account fails to consider the richness of the interaction afforded by dialogue. Although Schegloff’s sociological starting point (i.e., in terms of organizational practice and interaction contingen-
linguistic alignment at one level leads to linguistic alignment. It may lead to linguistic alignment, just as repeated action scenarios (feeding, bathing, etc.) may play an important role in the acquisition of syntax (Tomasello 2003). These suggestions help reinforce the claim that nonlinguistic alignment may lead to linguistic alignment, just as linguistic alignment at one level leads to linguistic alignment at other levels (see our discussion of Fussell & Kraut in sect. R7). In fact, Dominey suggests that such linguistic/nonlinguistic links are necessary to explain the process of language acquisition, where other participants (as in our discussion of Fussell & Kraut in sect. R7). In fact, Dominey suggests that such linguistic/nonlinguistic links are necessary to explain the process of language acquisition, where one partner (the infant) does not initially have linguistic abilities. A full theory of how interactive alignment might explain acquisition would be fascinating. In particular, we are intrigued by the suggestion that learning by alignment might avoid the enlistment of generative grammar mechanisms, perhaps in a way similar to that envisaged by Tomasello.

Language acquisition is a good example of how it may be possible to extend our account into new domains. Other areas that we have highlighted at various points in the target article and this response include social psychology and human-computer interaction. A recurring theme is that it may be possible to include nonlinguistic alignment into developments of our model; interlocutors who are aligned in nonlinguistic (e.g., body posture) or paralinguistic (e.g., tone of voice) ways may be more likely to align linguistically.

We emphasize that our use of the term “priming” is at a fairly abstract functional level, as our notion of automaticity makes clear (sect. R4). It allows nonconscious mediation by factors that may originate in distinctions that interlocutors are aware of (e.g., participant status, social status, cooperativeness). We also note that “priming” may employ transient activation or implicit learning or both. To be more speculative, we suspect that interactive alignment may work by two distinct mechanisms: a brief activation-based process that may not be affected by intentional distinctions, and a longer-lasting memory-based process that is intentionally mediated. The effects of these two processes will depend on precise timing, and will therefore be differentially affected by aspects of the conversation that affect timing. For example, a high-engagement face-to-face dialogue between intimate friends may result in timing that is precisely attuned to increasing alignment, whereas a dialogue between strangers that depend on external factors such as rules of engagement (e.g., in an interview) or technology (e.g., walkie-talkies) may not. We suspect that the longer-lasting process will not be affected but the activation process might be impaired in low-involvement dialogue. These speculative comments could inform an extensive program of empirical research concerned with the conditions that lead to alignment in dialogue (e.g., its time course).

The other obvious area for development is explicit computational modeling, as highlighted by Brown-Schmidt & Tanenhaus in particular. To perform such modeling, it would of course be necessary to explicate many assumptions of our account that are currently vague or implicit, for instance by developing interactive alignment, interactive repair, and other-modeling components. It would be necessary to model the process whereby alignment at one level leads to alignment at other levels, and to understand how conflicts of alignment are resolved (see Warren & Rayner). We need to know whether transient activation and implicit learning should be distinguished, and if so, how they interact. Finally, any such account should explain the process of routinization and describe its effects on alignment.

NOTE

1. Note that the uses of “interactive” in interactive alignment and interactive activation are unrelated.

References

Letters “a” and “r” appearing before authors’ initials refer to target article and response respectively.


References/Pickering and Garrod: Toward a mechanistic psychology of dialogue


References/Pickering and Garrod: Toward a mechanistic psychology of dialogue