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

Niels O. Schiller and Jan Peter de Ruiter

Some notes on priming, alignment, and self-monitoring

Niels O. Schiller and Jan Peter de Ruiter

*Department of Cognitive Neuroscience, Faculty of Psychology, Maastricht University, 6200 MD Maastricht, The Netherlands; ‡Max Planck Institute for Psycholinguistics, 6500 AH Nijmegen, The Netherlands.

JanPeter.deRuiter@mpi.nl Niels.Schiller@mpi.nl

Abstract: Any complete theory of speaking must take the dialogical function of language use into account. Pickering & Garrod (P&G) make some progress on this point. However, we question whether their interactive alignment model (IAM) is the optimal approach. In this commentary, we specifically critique (1) their notion of alignment being implemented through priming, and (2) their claim that self-monitoring can occur at all levels of linguistic representation.

The primary way of language use is dialogue, not monologue. We want to acknowledge the authors’ effort to stress this important point, which needs to be addressed explicitly in empirical and modeling work in speech production and comprehension research. We believe that these issues are especially relevant for syntactic processing. For instance, one wonders how syntactically incomplete (dialogue) utterances can be syntactically encoded in more traditional models, if there is no overt verb present in the generated utterance. Take, for example, the following extract from the dialogue transcript in section 2 of the target article:

1. A: . . . Tell me where you are?
   [Utterances 2 and 3 omitted]

2. A: Right: \{I am\} two along from the bottom one up.* [our addition in curly brackets]

In this example, speaker A does not produce the appropriate verb form of “to be” (i.e., “I am”) but nevertheless gives an acceptable and cooperative answer to speaker B’s question. This type of ellipsis can only be correctly produced if the syntax generator has access to previously stored discourse information, allowing the speaker to omit “I am,” even though the original question containing the verb occurred several utterances earlier in the discourse (see also Levelt 1989, p. 99, for a similar analysis).

Although we agree in principle with the authors’ assessment that the dialogical structure of language should receive more attention in accounts of language processing, we are not convinced that adopting the interactive alignment model is the right way to do so. For instance, it is unclear to us exactly how priming can account for alignment, and, in particular, we fail to see in what way priming is more than “a behavioral effect” (see target article, sect. 2.2). We believe that “priming” does not explain or implement interactive alignment. Real interactive alignment necessarily involves storing selected fragments from previous utterances. Priming can raise the probability of certain linguistic structures being selected, but this is not sufficient for the strong and explicit type of alignment the authors want to incorporate in models of language processing. Also, syntactic priming effects are weak effects. It is hard to see how an elaborate mechanism such as interactive alignment could be realized by only raising the probability of selecting a certain syntactic construct by roughly 10% to 20% (see, e.g., Pickering & Branigan 1998).

Our second critical note concerns one of the few testable predictions from the interactive alignment model, namely, that self-monitoring by the speaker occurs at all levels of linguistic representation. The authors explicitly claim that speakers can self-monitor abstract gender features (or any other syntactic features) at any level of linguistic representation that can be aligned (i.e., semantic, syntactic, lexical, phonological, and phonetic representations) – and not only at the phonological level.

For example, the authors explicitly claim that speakers can correct gender errors, such as *le tête instead of la tête* (“the head”) in French or *de been instead of het been* (“the leg”) in Dutch not only after they have been articulated but even before their overt production. This is an interesting claim that needs to be investigated in the future. However, we are somewhat skeptical about this claim because to our knowledge there is no evidence that self-monitoring of gender features (or any other syntactic features) is possible. For example, Desrochers and his collaborators (Desrochers & Paivio 1990; Desrochers et al. 1989; Muller-Gass et al. 2000) found that selecting a gender label (e.g., feminine or masculine) took about 200 msec longer than selecting the indefinite article in French gender decision. Furthermore, Tucker et al. (1977) provided empirical evidence suggesting that French speakers implicitly construct a noun phrase including the article and the noun to determine a noun’s gender. However, if speakers can self-monitor abstract gender information at the level of syntactic representation, as suggested by P&G, why would they go through the trouble of generating the gender-marked article of a noun to determine its gender?
In contrast to these findings about syntactic representations, recent evidence from our own laboratory as well as from other laboratories demonstrated that self-monitoring does occur at the level of phonological encoding. We have empirical data about the monitoring of phonological segments (Schiller, in press; Wheel- 
don & Morgan 2002), word stress (Schiller 2001; Schiller et al., in press), syllable boundaries (Jansma & Schiller 2004), and syllables (Morgan & Wheeldon 2003). However, we also have evi-
dence that participants are unlikely to monitor a phonetic-
acoustic representation of the respective utterances. Although gender decision as a task is widely used (Müller & Hagoort 2001; Schiller et al. 2003; Schmitt et al. 2001a; 2001b; Van Turenmout et al. 1998), it remains to be shown whether or not this task actually 
taps syntactic processing, because abstract gender informa-
tion may not be directly available to the speaker. Rather, gender information may be available only via its phonological realiz-
ation, for example, an article or gender-marking suffix. Interestingly, ef-ects of gender congruency have recently been re-interpreted as determiner congruency effects occurring at the phonological level and not at the gender feature level (see overview in Schiller & Caramazza 2003).

To conclude, we believe that internal self-monitoring does not occur at every single level of linguistic representation (as claimed by P&G) or at every processing level in models of speech produc-
tion (e.g., Levelt et al. 1999). Rather, we believe that internal evidence that internal self-monitoring works on phonological representa-
tions, which are created during phonological encoding in speak-
ing, for example, when segments are prosodified into phonologi-
cal words. It is at this level that information about segments, syllables, syllable boundaries, and word stress is available to the 
speaker. Although the abundance of evidence for phonological-
level monitoring does not necessarily exclude other monitoring levels (e.g., at the conceptual level; see Levelt 1989), we are un-
aware of any evidence for self-monitoring at earlier or later levels of grammatical and form encoding.

ACKNOWLEDGMENTS
Nieh O. Schiller is supported by a grant from the Netherlands Organiza-
tion for Scientific Research (NWO). Jan Peter de Ruiter is supported by 
the European Union Project Conversational Multi-modal Interaction with 
Computers (COMIC), grant IST-2001-32311. The authors thank Pim 
Levelt for his comments on an earlier draft.

Just how aligned are interlocutors’ representations?

Michael F. Schober
Department of Psychology, F330, New School for Social Research, New 
York, NY 10003. schober@newschool.edu
http://homepage.newschool.edu/~schober

Abstract: Conversational partners' representations may be less aligned 
when they do not understand each other, as data from a series of experiments on surveys about facts and be-

caracterizations pose (as proposed in Clark & Wilkes-Gibbs 1986) and yet never 
actually discover that their conceptions were misaligned. An 
important contention in the P&G article is that seemingly 
complex interactional processes can be modeled largely with sim-
ple individual mechanistic processes. The proposal is that conver-
sational partners, following a principle of parsimony, do not ordi-
narily model each other's mental states or make inferences about 
common ground, except when there is evidence that not doing so 
has led to obvious misunderstanding and when cognitive re-
sources allow.

I would argue that none of the current data actually allow us to 
distinguish this position from an alternative: that the ordinary case 
is that conversational partners do model each other, and that they 
fail to do so only when they are under heavy cognitive load or when 
circumstances weigh heavily against doing so. Why should we 
assume that the ordinary case is one where the interlocutor does not 
need to be modeled and the speaker is under heavy cognitive 
load? As far as I can tell, no one knows the level of load encoun-
tered in the range of ordinary interactive situations. As Susan 
Brennan and I have argued (Schober & Brennan 2003), the evi-
dence for egocentric processing is far from conclusive; the exper-
iments purported to show egocentric processing as basic rely on 
null results and experimental methods that are far removed from 
ordinary processing situations. When such studies are carried out 
in more realistic settings, the findings can look rather less ego-
centric.

Not to overstate the case, but one could argue that modeling 
ones’ partners only when it is needed may require a level of situa-
tional monitoring that leads to a paradox: How can one know ex-
actly when one needs to model one's partner without already