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; Wheeldon & 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 evidence 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 Turennout et al. 1998), it remains to be shown whether or not this task actually taps syntactic processing, because abstract gender information may not be directly available to the speaker. Rather, gender information may be available only via its phonological realization, for example, an article or gender-marking suffix. Interestingly, effects 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 production (e.g., Levelt et al. 1999). Rather, there is abundant evidence that internal self-monitoring works on phonological representations, which are created during phonological encoding in speaking, for example, when segments are prosodified into phonological 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 unaware of any evidence for self-monitoring at earlier or later levels of grammatical and form encoding.

ACKNOWLEDGMENTS

Niels O. Schiller is supported by a grant from the Netherlands Organization 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 than they appear even when interlocutors believe they have successfully understood each other, as data from a series of experiments on surveys about facts and behaviors suggest. Although the goal of a mechanistic psychology of dialogue is laudable, the ultimate model is likely to require far greater specification of individual and contextual variability.

When conversational partners come to agree that they are talking about the same thing, how aligned are their conceptualizations? The interactive alignment account holds that when dialogue is successful, interlocutors' linguistic representations are aligned at multiple levels. Although Pickering & Garrod (P&G) observe that alignment is sometimes only approximate and that evident misalignments can be interactively repaired, the general thrust of their approach is that successful communication requires representations to be the same.

I contend that interlocutors' using the same words can actually mask a surprising degree of undetected misalignment. Fred Conrad and I, with other colleagues, have carried out a series of laboratory and field studies examining how people interpret ordinary words in ongoing U.S. government surveys about facts and behaviors, words like "job," "bedroom," "smoking," and "cigarettes" (Conrad & Schober 2000; Schober & Conrad 1997; Schober et al. 2004; Suessbrick et al. 2000). Because the agencies that carry out these surveys have thorough definitions for the terms, answers to the survey questions provide evidence about the extent to which respondents' conceptions match the survey designers'. Our finding is that people's representations are frequently wildly misaligned with the survey designers' – and with each other's – without anyone's noticing.

For example, in one study (Suessbrick et al. 2000), survey respondents interpreted terms like "smoking" and "cigarettes" in a question such as "have you smoked at least one hundred cigarettes in your entire life?" differently enough (tobacco, cloves, marijuana? Finished or just a puff? Bought or borrowed?) that 10% of the respondents subsequently presented with a definition changed their answer to the question from yes to no or from no to yes. In a national telephone sample (Conrad & Schober 2000), more than 40% of reported purchases did not fit the survey designers' definitions, even though the questions had been widely pretested. And this is not just because the official definitions failed to match the population consensus about the meaning of terms; respondents' interpretations differed from each other's as much as they differed from the survey designers'.

Across our various studies, respondents are quite surprised at the thought that someone else might interpret the same words differently from the way they do; when given the opportunity to request clarification about the meanings of survey terms, they choose to do it a very small percentage of the time. People seem to follow a "presumption of interpretability" (Clark & Schober 1991): It should be the questioner's responsibility to forestall misinterpretation.

These data suggest a far more Quinian view of successful referring than the P&G account encompasses: Seemingly successful referring can mask conceptual misalignments that reflect deep underlying indeterminacies. The point is that people can believe they have understood each other well enough for current purposes (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 simple individual mechanistic processes. The proposal is that conversational partners, following a principle of parsimony, do not ordinarily 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 resources 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 encountered in the range of ordinary interactive situations. As Susan Brennan and I have argued (Schober & Brennan 2003), the evidence for egocentric processing is far from conclusive; the experiments 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 egocentric.

Not to overstate the case, but one could argue that modeling one's partner only when it is needed may require a level of situational monitoring that leads to a paradox: How can one know exactly when one needs to model one's partner without already

https://doi.org/10.1017/S0140525X04420056 Published online by Cambridge University Press

knowing what the partner model would clarify? (This seems akin to Fodor's frame problem or the notorious conundrums about selective attentional filtering). Arguably, it might be simpler to model the partner as the ordinary case. P&G assume that representing common ground requires unusually complex mental representations; this may be so, but it is also possible that representations of common ground are ordinary memory representations with the right content.

In fact, what is parsimonious may well vary across individuals and conversational contexts. Our notions of what is effortful in dialogue are not well worked out (Schober 1995; 1998a); people's knowledge of each other's perspectives, interest in modeling their partners, and abilities for doing so probably vary substantially more than current theories allow. For example, people who have very poor mental rotation abilities may be unable to conceive of a partner's point of view on a scene, leading to a different sort of partner modeling than that carried out by mental rotation whizzes (Schober 1998b). And different abilities may be relevant in different conversational contexts; mental rotation ability may have little to do with the consideration of lexical alternatives in dialogues that are not about physical scenes, whereas working memory capacity differences or different conversational agendas (Russell & Schober 1999) might. P&G are to be applauded for furthering the attempts to bridge the "language as product" and "language as action" traditions, but they have also opened a can of worms; the ultimate mechanistic account of dialogue is likely to require far greater specification of individual and contextual variability.

Dialogue processing: Automatic alignment or controlled understanding?

Hadas Shintel and Howard C. Nusbaum

Department of Psychology, The University of Chicago, Chicago, IL 60637.
hadas@uchicago.edu **hcn@speech.spc.uchicago.edu**

Abstract: Pickering & Garrod's (P&G's) mechanistic account of dialogue assumes that linguistic alignment between interlocutors takes place automatically, without using cognitive resources. However, even the most basic processes of speech perception depend on resource use. The lack of invariant mapping between input patterns and interpretations in dialogue, as in speech perception, may require controlled, rather than automatic, processing.

In their target article, Pickering & Garrod (P&G) challenge current psycholinguistic theory by proposing the study of dialogue as the appropriate paradigm for understanding language processing. They argue that the main problem facing interlocutors is the alignment of their respective situation models, and they propose a mechanistic account of alignment in dialogue, based on an interactive and resource-free priming mechanism. Linguistic structures, produced on the surface by one interlocutor, putatively prime corresponding structures in the other interlocutor. While we agree that conversation, rather than decontextualized, isolated sentences, should be taken as basic form of language use and as the foundation upon which language has evolved, it seems less plausible that much of the cognitive work in conversation takes place automatically.

P&G's model of automatic alignment assumes that a mapping can be established directly between the representations uttered by one interlocutor and those activated in the head of the other interlocutor. But as they themselves note, the linguistic information is encoded in sound. Hence, for the listener the spoken utterance must be recognized antecedent to any alignment occurring. However, even findings regarding the basic processes of speech perception argue against the assumption of automatic processing. Automaticity implies a passive process in which the input is processed in an invariant, inflexible manner, regardless of the beliefs and expectations of the listener (Nusbaum & Schwab 1986). For a process to be automatized, there must be a consistent mapping between input patterns and responses, the benefit of which is a process that does not impose demands on cognitive and attentional resources (Shiffrin & Schneider 1977).

Of course the hallmark of speech is the lack of consistent, invariant mapping between acoustic patterns and linguistic categories across phonetic contexts (Liberman et al. 1967) and across talkers (Peterson & Barney 1952). As a result of variable mapping, speech perception depends on controlled, active processing requiring attention and working memory (Nusbaum & Magnuson 1997). For example, variation between talkers (as in the circumstance of a conversation among three interlocutors) requires talker normalization, the process by which listeners compensate for acoustic-phonetic variability in vocal characteristics (see, e.g., Nearey 1989). Nusbaum and Morin (1992) found that talker variability slowed speech recognition (see also Mullennix & Pisoni 1990), and this slowing was because of increased demand on working memory.

Furthermore, talker normalization is contingent on listeners' expectations regarding the interpretation of acoustic patterns – which should not happen for an automatic direct-mapping process. Magnuson and Nusbaum (1994) demonstrated that when listeners expected a pitch difference to signal a talker difference, they showed talker normalization, but if the same pitch difference was expected as a way one talker accented speech, no normalization occurred. Similarly, the expectation that a talker was male or female significantly changed the interpretation of vowel tokens (Johnson et al. 1999). The acoustic patterns of speech are processed differently depending on listeners' expectations, arguing against invariant automatic processing. Expectation effects suggest that alignment may not be a result of a direct, automatic, causal link between the activation of a representation in one interlocutor and the activation of an objectively matching representation in the other interlocutor. Rather, the process is mediated by listeners' expectations. For example, alignment at the level of articulation, where interlocutors converge on each other's speaking style, may be mediated by interlocutors' stereotypic expectations about the other interlocutor's accent and speech rate, resulting in subjective but not objective alignment (Thakerar et al. 1982).

Listeners may need to use controlled active processing because the one-to-many mapping (one pattern may have multiple interpretations) in speech represents a nondeterministic computational problem that cannot be solved, in principle, by a deterministic system such as an automatic process (cf. Nusbaum & Magnuson 1997). This problem of one-to-many relationships between linguistic patterns and interpretations occurs across all levels of linguistic analysis (Nusbaum & Henly 1989). Indeed, research has shown that the same spoken sentence can be differently interpreted in different visual referential contexts (Tanenhaus et al. 1995) and that the same indirect requests can be processed differently as a result of speaker status (Holtgraves 1994). Although P&G focused on utterance-level matches that can occur within stretches of dialogue, real conversation is less predictable and routinized – otherwise communication would be unnecessary.

This raises several deep questions facing dialogue research: How do interlocutors cope with variability in linguistic and perceptual context? How are diverse sources of information integrated to constrain production and comprehension? And how do interlocutors flexibly adapt to different conversational circumstances? Although in real conversations these questions are often subjectively resolved quickly and without apparent effort, this is seldom an accurate barometer of the demands on cognitive resources such as attention and working memory.

P&G propose that the process of automatic alignment bypasses the need for modeling the interlocutor's mental states and for distinct, conscious decision stages. We agree that a constant monitoring of common ground would be unnecessary and costly, but the dichotomy between automatic processes and conscious decision processes involving complex inferences about the interlocutor's mental state does not represent the full range of possible pro-

https://doi.org/10.1017/S0140525X04420056 Published online by Cambridge University Press