ture. They provide a thought-provoking theory of dialogue in which coordination in message production occurs when interactants generate their messages from similar situation models and mimic their partner's production at the syntactic, semantic, lexical, phonological, and phonetic levels, based on primitive priming mechanisms. They argue that these alignment processes plus techniques for repairing misalignments are sufficient to explain most cases of what others have considered evidence of a deeper type of perspective-taking, in which speakers take their partners' mental states into account in forming their own speech.

We believe, however, that P&G's theory cannot completely explain recent data about language production. In our own work, for example, we find evidence across several experiments that when interactants in referential communication tasks have different views of a physical space, they accommodate their language to their partner's view rather than mimicking their partner's expressions (e.g., Fussell et al. 2000; 2003; Kraut et al. 2002; 2003). These data are consistent with the hypothesis that interactants are taking the perspective of their conversational partners.

Consider, for example, the case of deictic reference in a bicycle repair task (Kraut et al. 2003). In this task, one person (the "worker") performs a series of repair tasks under the guidance of a second person (the "helper"). The helper is located either beside the worker, where both can see and interact with the work area, or in a separate room connected only by an audio link. In a third condition, they are connected by an audio/video link through which the helper can see what the worker is doing but cannot interact with the work area. The conversations typically consist of helper's instructions followed by worker's actions, questions, or acknowledgments of understanding. Interactants can refer to task objects and locations with either extended linguistic expressions (e.g., "take the long dangling piece and put it in where the two large screws are") or shorter deictic references (e.g., "take *this* piece and put it *there*").

As Figure 1 shows, the ways in which workers refer to parts, tools, and other task objects depend on their partners' ability to see the work area. In the side-by-side condition, both helpers and workers can view one another and task objects, and both use a large number of deictic expressions. In the audio-only condition, the remote helpers cannot see the work space, and neither workers nor helpers use deictic expressions. The interesting case, from an alignment point of view, is the video condition. Here, the helpers can see the workers and work space but cannot point to objects in it. Under these conditions, helpers rarely use deixis. However, workers can point to task objects and they know that helpers can see them do so through the video link. They use deixis instead of matching the helpers' nondeictic expressions. If conversational alignment were driven by primitive priming mechanisms, then the workers should use nondeictic references in the video condition, after hearing helpers' uttering many of these expressions. (Because the helper could not be seen, he or she would have no way of using deictic expressions to match the worker's utterances.) In short, the way workers referred to task objects and locations depended upon what their partners could see, not the language their partners previously used to refer to these same objects and locations.

We believe these results demonstrate that one type of deep common ground – visual copresence – is assessed during message production, at least for lexical selection processes. Indeed, in experiments where the views can change, interactants often explicitly exchange information about what each can see, with phrases such as: "Can you see the table?" (Kraut et al. 2002; 2003).

P&G might argue that video-mediated communication is a nonprototypical dialogue setting and hence may elicit special processes of assessing deep common ground. Note, however, that the audio-only discourse from Garrod and Anderson's (1987) maze study is similarly nonprototypical. Rather than demonstrating that people in face-to-face dialogues use processes of verbal alignment in lieu of deeper considerations of common ground, Garrod and Anderson's results may indicate that people verbally align primarily when the context has been stripped of all other indicators of common ground.

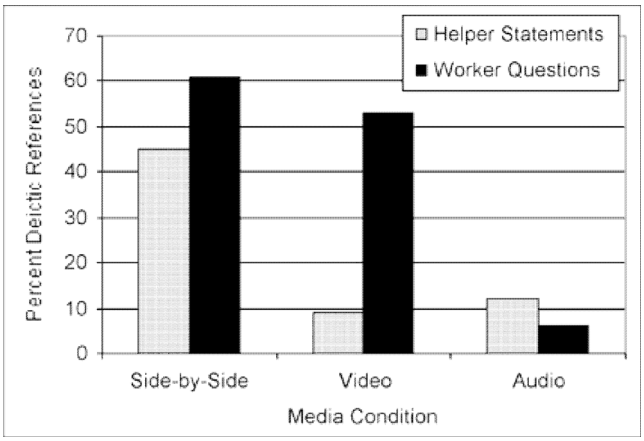


Figure 1 (Fussell & Kraut). Percentage of references to task objects containing verbal deixis, by media condition and participant role (from Kraut et al. 2003).

In their discussion of deep common ground versus automatic alignment, P&G in essence take a straw-man approach to describing the processes involved in conversational grounding. As Clark and Marshall (1981) have discussed, common ground can be determined using heuristics based on community comembership, linguistic copresence, and physical copresence. Some calculations of common ground (e.g., a helper trying to determine whether a worker on the bicycle task knows what a derailleur is) may be difficult; others (e.g., a worker trying to determine whether the helper can see the work space) may be relatively easy. Ruling out deep common ground as a fundamental process in dialogue production would require a series of carefully controlled studies that have not been performed to date. P&G's article is valuable for the detail with which it specifies an alternative model that would need to be included in such experiments.

We conclude by observing that calculations of deep common ground are essential for determining when to speak and what to say. For example, in the bicycle repair studies, in the audio condition, when the helpers cannot see them, workers describe what they are doing. Helpers rely on these verbal reports to determine when to provide new instructions or clarify preceding ones. In the video and side-by-side conditions, workers do not bother describing what they are doing because they know that the helpers are watching their activities. If common ground is available to communicators for these processes of message timing and content, it should not require much additional effort for them to incorporate it into the messages themselves.

Intrinsic misalignment in dialogue: Why there is no unique context in a conversation

Jonathan Ginzburg

Dept of Computer Science, King's College London, London WC2R 2LS, United Kingdom. **ginzburg@dcs.kcl.ac.uk**
http://www.dcs.kcl.ac.uk/staff/ginzburg

Abstract: Pickering & Garrod's (P&G's) claim that conversationalists do not explicitly keep track of their interlocuters' information states is important. Nonetheless, via alignment, they seem to create a virtually symmetrical view of the information states of speaker and addressee – a key component of their accounts of collaborative utterances and of self-monitoring. As I show, there is significant evidence for intrinsic contextual misalignment between conversationalists that can persist across turns.

https://doi.org/10.1017/S0140525X04300051 Published online by Cambridge University Press

I am very sympathetic with Pickering & Garrod's (P&G's) central message: namely, that there is a need to develop a detailed, computationally oriented account of the mechanisms underlying language use in dialogue, and moreover, that these – not mechanisms developed for monologue – should be regarded as the primary mechanisms of the language faculty.[1] On a more technical front, I think that a key claim of the authors, that dialogue participants do not explicitly keep track of their interlocuters' information states, but rather, that this is emergent from the dynamic alignment of each other's information states, is an important one.

Nonetheless, in seeking intrinsically to couple dialogue participants via alignment, they seem to create a virtually symmetrical view of the information states of speaker and addressee. Thus, a key component of P&G's accounts of collaborative utterances as well of self-monitoring is the claim that who is speaking at a given point does not, in some sense, make a difference given alignment: The addressee can take over or the speaker can "change voice" and self-correct. As I show in this commentary, this claim is incorrect: There is significant evidence that the contexts available to the conversationalists are *not* identical. Hence, there is actually intrinsic contextual *misalignment* between conversationalists that can persist across turns.

That a common context (cf. the *common ground* prominent in work by Clark & Marshall 1981; Lewis 1979; Stalnaker 1978) *emerges* in dialogue is an important insight. It yields a better picture of, for instance, querying, than classical speech-act views provide (Searle 1969). So in asking a question, to take one example, a speaker puts up a question for discussion and whoever takes over the turn can address it – either the original asker or the original addressee:

1. A: Whom should we invite to the conference?
   A or B: Would Phil be a good idea?

And yet, the actual situation is not as symmetrical as this: The speaker's options for self-repair or, indeed, other follow-up are quite distinct from the addressee's options. This can be illustrated succinctly by a phenomenon I have dubbed the Turn Taking Puzzle (Ginzburg 1997a; 1997b). Questions of the form "Why?" involve radical context dependence – pre-theoretically, the context supplies a propositional referent of some kind (Moore 1995). Interestingly, examples 2a and 2b (below) show that the resolution accorded to the bare "why" changes according to who keeps or takes over the turn. The resolution that can be associated with "Why?" if A keeps the turn is unavailable to B if he or she had taken over, and vice versa:

2a. A: Which members of our team own a parakeet?
    A: Why? (= Why own a parakeet?)
2b. A: Which members of our team own a parakeet?
    B: Why? (= Why are you asking which members of our team own a parakeet?)
2c. Which members of our team own a parakeet? Why am I asking this question?

Example 2c shows that these facts cannot be reduced to coherence or plausibility – the resolution unavailable to A in example 2a yields a coherent follow-up to A's initial query if it is expressed by means of a nonelliptical form. In other words, the context is responsible for these interpretational asymmetries or, rather, they are a consequence of the fact that distinct contexts are associated with the conversationalists.

Similarly, a common strategy for requesting a clarification is by means of a *reprise fragment* – a word or constituent of the previous utterance (see Purver et al. 2002; 2003 for corpus and experimental evidence on clarification requests, particularly reprise fragments). Reprise fragments have two prominent understandings (Ginzburg & Cooper 2004), exemplified in 3a (below). However, it is quite strange for a speaker to follow up her utterance with a reprise fragment. This becomes felicitous only if followed up by an additional correction, such as "Wait, did I say Bo, no I mean Lou," or some such. However, even then the readings that arise in example 3a, whose resolution is radically context-dependent, are not manifested:

3a. A: Did Bo leave?
    B: Bo? (= Either: Are you asking if BO of all people left?
    Or: Who were you referring to as "Bo"?)
3b. A: Did Bo leave?
    A: #Bo?[2]

It is worth noting that contextual asymmetries of this kind can persist for quite a number of turns – essentially as long as a given discourse topic remains under discussion. Example 4 is an extract from the British National Corpus in which Chris's "Why?" is naturally understood to refer to Norrine's utterance five turns back, an utterance which seems to be viewed as *grounded* (Clark 1996):

4. Norrine (1): When is the barbecue, the twentieth? (pause) Something of June.
   Chris (2): Thirtieth.
   Norrine (3): A Sunday.
   Chris (4): Sunday.
   Norrine (5): Mm.
   Chris (6): Why? (= Why do you ask when is the barbecue?)
   Norrine (7): Because I forgot (pause) That was the day I was thinking of having a proper lunch party but I won't do it if you're going out.

Note that the resolution associated with Chris's "Why?" is simply unavailable to Norrine at all subsequent points, as illustrated in the constructed variant of statement (4) in example 5a. As with previous examples, this cannot be explained on "pragmatic" grounds because the speaker can fairly coherently express the requisite reading in nonelliptical fashion, as in example 5b:

5a. Norrine (1): When is the barbecue, the twentieth? (pause) Something of June.
    Chris (2): Thirtieth.
    Norrine (3): A Sunday.
    Chris (4): Sunday.
    Norrine (5): # Why? (Cannot mean: Why is Norrine asking when is the barbecue?)
5b. Norrine (5′): Mm. Why am I asking? Can you guess?
    Chris (6′): No idea.

Phenomena such as this suggest that the different roles conversationalists play with respect to a given utterance (speaker vs. addressee) are not something that gets neutralized in the utterance's aftermath. The contextual possibilities available for one conversationalist differ from those of the other conversationalist (I am referring to dialogue here; as P&G point out, multilogue is a genre with various distinct properties from two-person dialogue). In other words, a single context is not fully adequate to describe dialogue, even when talking about "public" context, which results from overtly registered conversational actions. Instead, one needs to view dialogue as involving updates by each conversationalist of some type of a publicly accessible domain which is relative to each conversationalist and so is parametrizable by unpublicized factors such as individual goals and intentions (cf. Hamblin's *individual commitment slate,* Hamblin 1970). The framework named KOS spells out this view and develops theoretical accounts as well as computational implementations of illocutionary and metacommunicative acts, including a detailed account of puzzles such as the Turn Taking Puzzles exemplified above (cf. Cooper et al. 2000; Ginzburg 1996; 2002; forthcoming; Larsson 2002).

ACKNOWLEDGMENTS

The research described here is funded by grant number RES-000-23-0065 from the Economic and Social Research Council of the United Kingdom and by grant number GR/R04942/01 from the Engineering and Physical Sciences Research Council of the United Kingdom.

NOTES

1. The authors point out the dearth of work in mechanistic psychology and theoretical linguistics (primarily by syntacticians) on dialogue. Since the late 1990s, however, there has been work by formal and computational

https://doi.org/10.1017/S0140525X04300051 Published online by Cambridge University Press

semanticists precisely on developing theories of information states and their dynamics in dialogue – see, for example, work within the EU TRINDI project (Consortium 2000) and the annual SEMDIAL series of conferences on the semantics and pragmatics of dialogue. Links to all previous conferences, starting in 1997, in Munich, are available from: http://cswww.essex.ac.uk/semdial/

2. The # symbol is used here, as standard in linguistics, to mark an infelicitous utterance.

Dialogue: Can two be cheaper than one?

Sam Glucksberg

Psychology Department, Princeton University, Princeton, NJ 08544.
samg@princeton.edu **http://www.princeton.edu**

Abstract: Pickering & Garrod (P&G) argue that language processing in dialogue is in principle easier than in monologue. Although dialogue situations may provide more opportunities for facilitative priming, those priming mechanisms are also available in monologue situations. In both cases, the interactive alignment model calls strict modular accounts of language processing into serious question.

Pickering & Garrod's (P&G's) argument for the primacy of dialogue is a refreshing alternative to the isolated sentence as the unit of analysis in psycholinguistics. Others, such as Krauss, Clark, and their students and colleagues, have made similar arguments (cf. Clark 1996; Fussell & Krauss 1992), but none have made the additional argument that much of dialogue production and comprehension is automatic, requiring minimal cognitive or linguistic resources. The details of the priming processes have yet to be worked out in detail, but P&G provide enough empirical evidence to warrant optimism about an eventual fleshing out of their interactive alignment proposal. Beyond such details, their account has several intriguing implications, not the least of which is the claim that language production and comprehension are easier in interactive dialogue than in monologue. This flies in the face of the consensus that dialogue requires complex inferences about the mental state of one's interlocutor that are not required when one is, say, talking to oneself, or processing utterances under monologue conditions.

Is language processing in dialogues inevitably easier (more fluent, requires less processing, etc.) than in monologues? I do not think this question can be answered unequivocally. I can think of contexts in which dialogue language production would be quite difficult and contexts in which monologue production would be quite easy. An example of the former might be trying to hold a conversation with a sullen teenaged son or daughter on the merits of atonal music. Examples of the latter might include the think-aloud paradigm used in research on reasoning and problem-solving, or young children talking to themselves or to their dolls when no one else is around. So, as usual, the answer is: It depends.

A more useful question addresses the ways in which dialogue might be more likely than monologue to facilitate language processing. Such ways might in principle be available only in dialogue. Alternatively, it might just be a matter of likelihood, rather than a principled difference between monologue and dialogue. Consider the stages of speech production that could be facilitated in monologues and in dialogues. Following Levelt (1989), production could be facilitated at any or all of the following stages: the message (what to talk about); grammatical encoding (syntactic representation and lexical access); phonological encoding; and articulation. One of the more obvious ways to facilitate speech production is to provide something to talk about. In dialogue, the situation (including one's interlocutor) usually provides this, but of course there are (sometimes painful) situations where one does not know what to talk about next. In monologue, including the occasional classroom lecture (or a *BBS* commentary, for that matter), one can be at a loss as to what to say or write, but this is not a principled characteristic of monologues, just an unfortunate occasional one.

Grammatical and phonological encoding, on the other hand, may well be advantaged in dialogue via the mechanisms of priming and the opportunities for imitation. P&G make a convincing argument that such priming not only occurs but may well be automatic. However, such priming could in principle operate in monologue as well, although the opportunities are undoubtedly less likely. In monologue, as in dialogue, opportunities to repeat specific syntactic forms should facilitate successive productions. Similarly, opportunities to repeat specific word senses should facilitate lexical access and sense selection. What's missing from monologue, of course, are the contributions of an interlocutor, which can provide opportunities for imitation, as well as for syntactic, lexical, and phonological priming (but see Ferreira & Griffin 2003 for evidence of potential sources of priming in monologue).

Finally, P&G note that contextual constraints might operate more efficiently in dialogue than in monologue, but, again, this is probably not a principled difference. They write "With respect to lexical ambiguity, we predict that context will have a very strong role, so that effects of meaning frequency can be overridden" (sect. 5.4, para. 3). To my knowledge, this prediction has not yet been tested under dialogue conditions. However, it has been confirmed under monologue conditions, most recently by Sereno et al (2003), who found very early context effects on word recognition that overrode frequency effects.

To the extent that automatic priming can facilitate language comprehension and production, dialogue may well afford more opportunities for such priming than monologue. Nevertheless, the same mechanisms should be available for both. P&G's contribution is thus more profound than the observation that language processing in dialogue might be easier than in monologue. Instead, the contribution is in the insight that there may be mechanisms of language production and comprehension that can circumvent the need for costly cognitive processing in interactive conversation. People may not need to use a theory of mind or make inferences about common ground and mutual knowledge. They may, instead, rely on automatic priming both within and between interlocutors to produce alignment of representations, which, in turn, makes for successful and relatively effortless speech processing. The theoretical implication for theories of language processing is clear: A strictly modular, bottom-up language processing model simply will not do the job.

ACKNOWLEDGMENT

I thank Boaz Keysar for helpful comments on an earlier draft of this commentary.

Resonance within and between linguistic beings

Stephen D. Goldinger[a] and Tamiko Azuma[b]

[a]*Department of Psychology, Arizona State University, Tempe, AZ 85287-1104;* [b]*Department of Speech and Hearing Science, Arizona State University, Tempe, AZ 85287-0102.* **goldinger@asu.edu** **azuma@asu.edu**

Abstract: Pickering & Garrod (P&G) deserve appreciation for their cogent argument that dialogue merits greater scientific consideration. Current models make little contact with behaviors of dialogue, motivating the interactive alignment theory. However, the theory is not truly "mechanistic." A full account requires both representations and *processes* bringing those representations into harmony. We suggest that Grossberg's (1980) adaptive resonance theory may naturally conform to the principles of dialogue.

https://doi.org/10.1017/S0140525X04300051 Published online by Cambridge University Press