Fussell, S. R., & Kraut, R. E. (in press).
Visual co-presence and conversational coordination

Commentary on Pickering & Garrod (In press).
Toward a mechanistic psychology of dialogue
Behavior and Brain Sciences.

ABSTRACT

Pickering and Garrod's theory of dialog production cannot completely explain recent data showing that when interactants in referential communication tasks have different views of a physical space, they accommodate their language to their partners' view rather than mimicking their partner's expression. Instead, these data are consistent with the hypothesis that interactants are taking the perspective of their conversational partners.

FULL TEXT

We applaud Pickering and Garrod's attempt to explain one of the most basic features of human language—its dialog structure. 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 Pickering and Garrod's theory of dialog production 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 partners' view rather than mimicking their partner's expressions (e.g., Fussell, Kraut, & Siegel, 2000; Fussell, Setlock, & Kraut, 2003; Kraut, Fussell, & Siegel, 2003; Kraut, Gergle, & Fussell, 2002). 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 either located 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 acknowledgements of understanding. Interactants can refer to task objects and locations either with 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 depends on their partner's 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 helper can't see the workspace, and neither uses deictic expressions. The interesting case, from an alignment point of view, is the video condition. Here, the helper can see the worker and workspace but can't 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' non-deictic expressions. If conversational alignment were driven by primitive priming mechanisms, then the workers should use non-deictic references in the video condition, after hearing helpers' uttering many of these expressions. (Because the helper could not be seen, he/she would have no way of using deictic expressions to match the workers utterances.) In short, the way workers referred to task objects and locations depended upon what their partner could see, not the language their partner previously used referred to these same objects and locations.

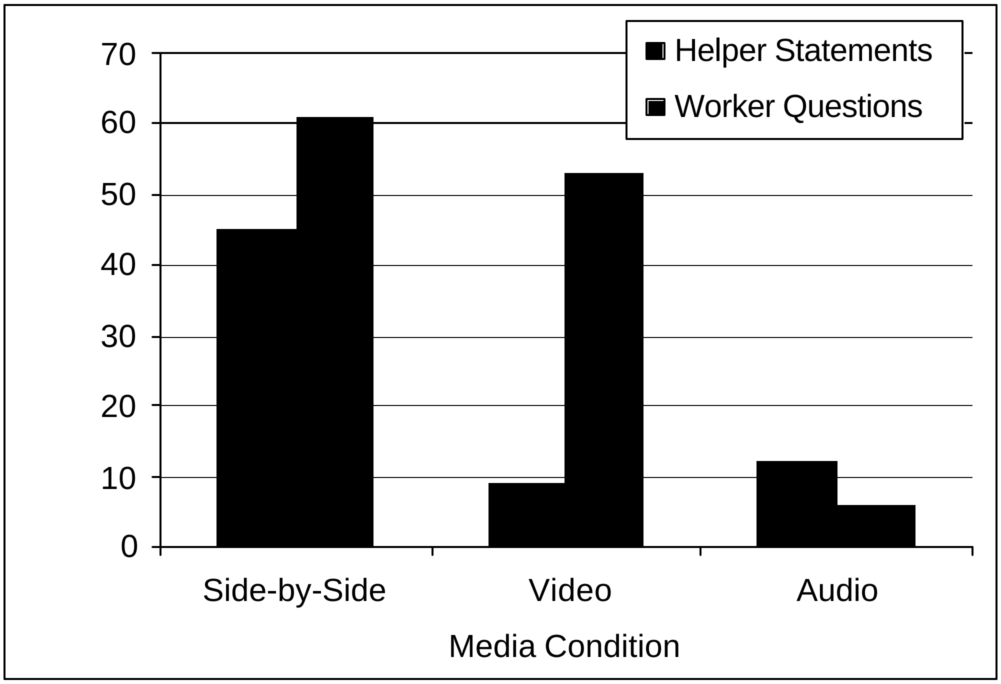

Figure 1. Percentage of deictic references to task objects by media condition and participant role (from Kraut, Fussell, & Siegel, 2003).

We believe these results demonstrate that one type of deep common ground—visual co-presence—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, Fussell, & Siegel, 2003; Kraut; Gergle, & Fussell, 2002).

Pickering and Garrod might argue that video-mediated communication is a non-prototypical dialogue setting and thus 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 non-prototypical. 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.

In their discussion of deep common ground versus automatic alignment, Pickering and Garrod in essence take a straw man approach to describing the processes involved in conversational grounding. As Clark and Marshall (1981) discussed, common ground can be determined using heuristics based on community co-membership, linguistic co-presence, and physical co-presence. 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 workspace) 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. Pickering and Garrod's paper 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, workers describe what they are doing in the audio condition, when the helpers can't see them. 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 don't bother describing what they are doing because they know that the helper is watching their activities. If deep 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.

References

Clark, H. & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*, 1-39

Clark, H. H. & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, R. M. Levine, & S. D. Teasley (Eds.). *Perspectives on socially shared cognition* (pp. 127-149). Washington, DC: APA

Clark, H. H. & Marshall, C. E. (1981). Definite reference and mutual knowledge. In A. K. Joshi, B. L. Webber & I. A. Sag (Eds.), *Elements of discourse understanding* (pp. 10-63). Cambridge: Cambridge University Press.

Clark, H. H. (1996). *Using language*. Cambridge, UK: Cambridge University Press.

Fussell, S. R., & Krauss, R. M. (1992). Coordination of knowledge in communication: Effects of speakers' assumptions about what others know. *Journal of Personality and Social Psychology, 62,* 378-391.

Fussell, S. R., Kraut, R. E., & Siegel, J. (2000). Coordination of communication: Effects of shared visual context on collaborative work. *Proceedings of CSCW 2000* (pp. 21-30). NY: ACM Press.

Fussell, S. R., Setlock, L. D., & Kraut, R. E. (2003). Effects of head-mounted and scene-oriented video systems on remote collaboration on physical tasks. *Proceedings of CHI 2003.*

Garrod, S., & Anderson, A.. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition, 27,* 181-218.

Kraut, R. E., Fussell, S. R., & Siegel, J. (2003). Visual information as a conversational resource in collaborative physical tasks. *Human-Computer Interaction, 18,* 13-49.

Kraut, R. E., Gergle, D., & Fussell, S. R. (2002). The use of visual information in shared visual spaces: Informing the development of virtual co-presence. *Proceedings of CSCW 2002.* NY: ACM Press.