Research Report

Using Pointing and Describing to Achieve Joint Focus of Attention in Dialogue

Adrian Bangerter

Stanford University

ABSTRACT—*Pointing was shown to focus attention in dialogue. Pairs of people talked and gestured to identify targets from arrays visible to both of them. Arrays were located at five distances: arm length (0 cm), 25 cm, 50 cm, 75 cm, and 100 cm. Some pairs could point; others could not. People relied more on pointing and less on language as distance decreased. Pointing especially suppressed descriptions of target location, suggesting that it was used to focus attention on a spatial region.*

The ease of dialogue belies the close coordination of participants' actions that takes place. People interacting coordinate turns (Sacks, Schegloff, & Jefferson, 1974), eye gaze (Argyle & Cook, 1976), and other behaviors. But one of the most important things they coordinate is attention. People have a *joint focus of attention* if they are attending to the same object and are mutually aware of it (Baron-Cohen, 1995; Clark & Marshall, 1981).

Joint attention simplifies referring by circumscribing a subdomain, either within discourse (Brennan, 1995; Grosz & Sidner, 1986) or within shared visual space. In a collaborative building task (Beun & Cremers, 1998), pairs used assumptions about joint visual attention to reduce collaborative effort. Speakers produced referring utterances (e.g., "the red block") that were ambiguous with respect to the task domain, but the utterances were effective because they were unambiguous with respect to the subdomain within attentional focus. Eye-tracking studies show that people use joint-attention assumptions to rapidly circumscribe subdomains (Brown-Schmidt, Campana, & Tanenhaus, 2002; Velichkovsky, 1995). Joint attention is an important conversational resource.

The present study investigated whether pointing is used to coordinate attention. Several authors have suggested it does (Buchler, 1940; Clark, 2003), but no experiments have directly addressed this question. Pointing gestures are closely coordinated with language (Marslen-Wilson, Levy, & Tyler, 1982) and processed automatically (Langton & Bruce, 2000). Inaccuracy in detection of the referents of pointing (Butterworth & Itakura, 2000) suggests that pointing shifts attention into the visual periphery, rather than identifying referents. In conversation, pointing marks initial reference to objects (Levy & McNeill, 1992) or new information (Clark, Van Der Wege, & Katz, 2002), further evidence of an attention-shifting function.

But pointing is often considered ambiguous (Pechmann & Deutsch, 1982; Schmidt, 1999), on the basis of the assumption that it is used mainly to locate a referent precisely. Standard accounts of deixis (Lyons, 1981) make this assumption. And it is embodied in study designs that involve selecting targets from a small number of alternatives (O'Neill & Topolovec, 2001; Thompson & Massaro, 1986). Such designs preclude the study of pointing as a device for achieving joint attention.

An alternative is that pointing is neither more nor less ambiguous than language. Rather, it is part of a composite signal combining both linguistic and gestural methods of reference (Bavelas & Chovil, 2000; Clark, 1996; Engle, 1998; McNeill, 1985; Schmauks, 1991). Different methods have different purposes. Pointing gestures circumscribe a referential domain by directing gaze to an approximate spatial region. Participants then assume joint attention and use reduced verbal methods. Of course, under certain circumstances (e.g., the referent is near), pointing can identify a target without language. In the composite-signal view, referring is opportunistic (Clark, 1996): Speakers minimize collaborative effort by trading off between linguistic and gestural methods, so that the relative burden on each varies according to the situation.

Three hypotheses were tested in the study reported here. The first was that the relative use of pointing and language varies according to the situation. As pointing becomes ambiguous, speakers will rely on it less and compensate with language. The second was that pointing is not redundant with speech. It reduces verbal effort to identify a target. The third hypothesis was that pointing focuses attention by directing gaze to the target region.

Pairs of participants talked and gestured freely to identify targets (photos of faces) from arrays visible to both participants. One participant, the director, identified each target to the other, the matcher. Ambiguity of pointing was operationalized by varying the distance of the arrays. Some pairs could see each other and therefore use pointing

Address correspondence to Adrian Bangerter, Groupe de Psychologie Appliquée, Université de Neuchâtel, Faubourg de l'Hôpital 106, 2000 Neuchâtel, Switzerland; e-mail: adrian.bangerter@unine.ch.

(*visible* pairs), whereas others could not (*hidden* pairs). The first hypothesis was tested by analyzing frequency of pointing as a function of distance in the visible condition. The second hypothesis was tested by examining the effects of condition and distance on verbal effort.

The third hypothesis was tested by assessing use of different verbal methods of referring: *feature* descriptions, which identified a target by its attributes ("the girl *with red hair*"); *location* descriptions, which described the position of the target in the array ("on *the far right*"); and *deictic* expressions ("*that* guy"), which accompanied pointing (Bühler, 1965; Lyons, 1981). Different methods contribute differently to referring. Feature descriptions contrast a target with competitors and, in principle, allow unambiguous identification (Olson, 1970). However, when there are many competitors, feature descriptions used in isolation may not minimize collaborative effort (Grice, 1975). If speakers also attempt to focus attention, they should use location descriptions before feature descriptions. By hypothesis, if pointing focuses attention, then it should sometimes be used in place of location descriptions.

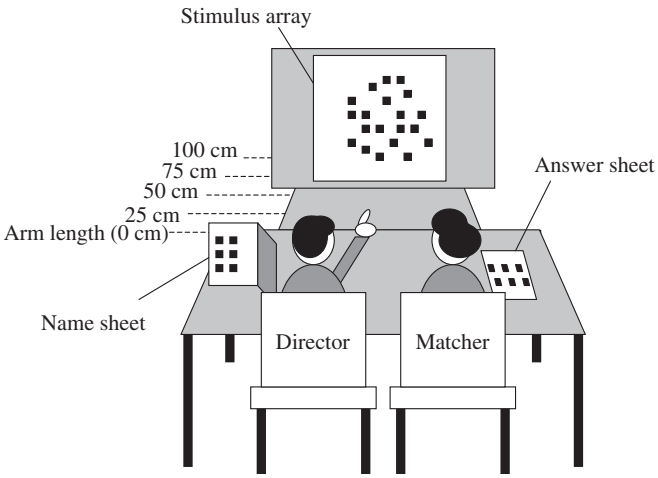

Fig. 1. Experimental setup.

METHOD

Participants

Forty Stanford University students participated in pairs (10 pairs per condition) for credit or pay. Participants were native English speakers with normal or corrected-to-normal vision.

Materials

The materials were five arrays, each consisting of twenty 4.5- × 4.5-cm color photos of faces. Each array was printed onto a 61- × 61-cm sheet and affixed to a large board for ease of handling. Arrays had 10 targets and 10 distractors each. No photo was used twice. Arrays 1, 3, and 5 had photos of women, and Arrays 2 and 4 had photos of men. Photos were arranged so as not to form obvious rows and columns.

Directors received name sheets indicating the 10 targets for each array. Each target's name was printed above the corresponding photo. Name sheets were hidden from the matchers' view. For each array, matchers received an answer sheet with the 20 photos of targets and distractors. There was space above each photo to write the person's name.

Procedure

The participants in each pair sat next to one another at a table on which the name sheets and the answer sheets were placed (Fig. 1). The person who sat on the left was the director, and the person who sat on the right was the matcher. A second table was placed to the side of the table opposite the participants, touching it in the middle and forming a T shape. An easel was placed on this table to keep the array boards vertical. Five distances were premarked for accurate placement of the easel: arm length (0 cm), 25 cm, 50 cm, 75 cm, and 100 cm. At arm length, the easel was flush with the side of the participants' table opposite them. They could touch the array by leaning forward.

In the hidden condition, a screen placed between the participants completely hid them from each other. It did not hide participants' views of the arrays at any distance.

Participants seated themselves in either chair, thus determining their roles. At the start of the session, the easel was positioned at the first distance (distances were presented in an essentially random order). The experimenter explained the task and instructed participants to proceed to identify the targets in the order on the director's sheet. They were told they could identify the pictures in any way they wanted and both talk as much as they wanted, but should remain seated. The experimenter placed the first array on the easel, started video recording, and told the participants to begin. The director began by trying to identify the first target and communicate that person's name to the matcher (e.g., "First is Ellen, at the top with glasses"). The matcher wrote the name down on the answer sheet, whereupon the director continued with the next target.

When all 10 targets had been identified, the experimenter moved the easel to the next distance, placed the second array on it, and instructed the participants to continue. This procedure was repeated until all five arrays were completed.

Data Collection and Coding

Two digital video cameras recorded the interaction, affording two views of director and matcher. Recordings were mixed onto a split-screen image for simultaneous viewing. Dialogue was transcribed word for word. For each target, coders noted whether or not (i.e., presence or absence) verbal (location, feature, and deictic descriptions) and gestural (pointing) methods were used by either person.

Coding of deixis was restricted to demonstrative pronouns or adverbs (*this*, *that*, *here*, *there*). Pointing was coded from the video and considered to be present when the arm was fully extended. Hidden pairs almost never pointed in such a way. When speaking, pairs in both conditions sometimes pointed with their index finger only (elbow resting on the table). Such points were not intended to be communicative and were disregarded in the present study. Reliability (Cohen's kappa), assessed by independent coding of 20% of the data, was acceptable to good (Fleiss, 1981): .76 for location descriptions, .73 for feature descriptions, .70 for deictic descriptions, and .65 for pointing, all $ps < .001$.

Visibility and distance constituted a 2 × 5 mixed-model design. Dependent variables were verbal effort (number of words per array used by director and matcher) and number of targets per array for which various referring methods were used.

RESULTS AND DISCUSSION

Pointing and Verbal Effort as a Function of Distance

Visible pairs pointed in referring to 52% of the targets. Thirty-two percent of points were accompanied by deixis (e.g., pointing and saying, "*That's* John"). Analysis revealed that pointing had different functions depending on whether or not it was associated with verbal deixis. Thus, points with deixis were analyzed separately from other points.

Pairs used more points with deixis as targets got closer (Fig. 2); in a repeated measures analysis of variance (ANOVA), the effect of distance (five levels, within subjects) was significant, $F(4, 36) = 17.9$, $p < .001$. Post hoc tests (Tukey HSD) showed that pairs used more points with deixis at arm length than at other distances. Their use of points without deixis did not vary with distance (Fig. 2), $F(4, 36) = 1.1$, n.s.

Pointing reduced verbal effort. Visible pairs used fewer words as targets got closer, whereas hidden pairs used the same number of words irrespective of distance (Fig. 3). A 2 (condition) × 5 (distance) ANOVA revealed a significant main effect of condition, $F(1, 18) = 12.1$, $p < .01$; a significant main effect of distance, $F(4, 72) = 7.2$, $p < .001$; and a significant interaction, $F(4, 72) = 6.5$, $p < .001$. Post hoc tests showed that visible pairs used fewer words at arm length than at 25 cm. They also tended to use fewer words at 25 cm than at 50 cm (but $p < .10$).

The two types of pointing affected verbal effort differently. The number of words used per array correlated negatively with the number of points with deixis ($r = -.62$, $n = 50$, $p < .001$), but was unrelated to the number of other points ($r = -.14$, n.s.).

Results thus support the first two hypotheses. The relative use of points with deixis and language varied with distance. Points with deixis reduced verbal effort. Other points were used independently of distance and did not reduce verbal effort.

Impact of Pointing on Location and Feature Descriptions

Visibility suppressed location descriptions. Together with the results already summarized, this supported the third hypothesis. Hidden pairs used location descriptions 99% of the time, irrespective of distance. Visible pairs used fewer location descriptions for closer than for farther targets (Fig. 2). A 2 (condition) × 5 (distance) ANOVA revealed a significant main effect of condition, $F(1, 18) = 30.3$, $p < .001$; a significant main effect of distance, $F(4, 72) = 26.3$, $p < .001$; and a significant interaction, $F(4, 72) = 25.7$, $p < .001$. Post hoc tests showed that visible pairs used location descriptions less at arm length (43% of the time) than at 25 cm (84%), and less at 25 cm than at 50 to 100 cm (97%).

Availability of gesture also suppressed feature descriptions, suggesting that pointing was also used to identify targets. Hidden pairs used feature descriptions 99% of the time, irrespective of distance. Visible pairs used fewer feature descriptions for closer than for farther targets (Fig. 2). A 2 (condition) × 5 (distance) ANOVA revealed a significant main effect of condition, $F(1, 18) = 13.5$, $p < .01$; a significant main effect of distance, $F(4, 72) = 6.1$, $p < .001$; and a significant interaction, $F(4, 72) = 6.0$, $p < .001$. Post hoc tests showed that visible pairs used feature descriptions less at arm length (66% of the time) than at other distances (93%).

Pairs in both conditions focused attention. Hidden pairs focused attention with location descriptions before using feature descriptions to identify the target: In these pairs, location descriptions preceded feature descriptions 91% of the time (chance: 50%), $t(9) = 11.2$, $p < .001$.

For visible pairs, the suppression of location descriptions was due to use of points with deixis. Pairs used location descriptions 92% of the time with other points. But with points with deixis, the rate of location descriptions dropped to 46%, $\chi^2(1, N = 499) = 107$, $p < .001$. The suppression of feature descriptions among visible pairs was also

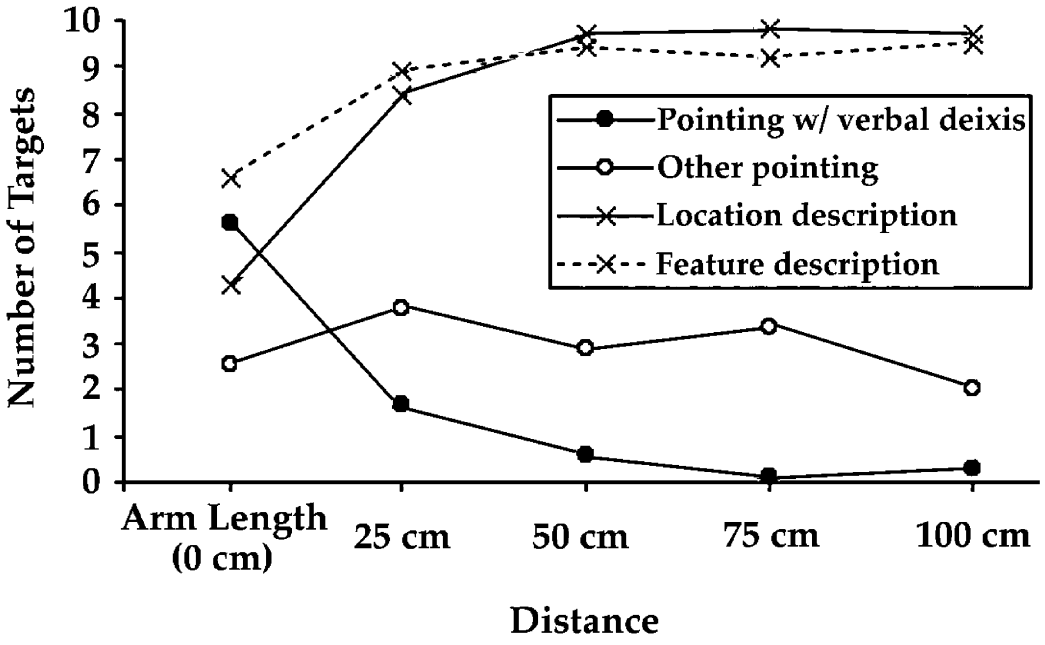

Fig. 2. Results from the visible condition: mean number of targets per array identified by pointing gestures accompanied by deictic expressions, other pointing gestures, location descriptions, and feature descriptions as a function of distance.

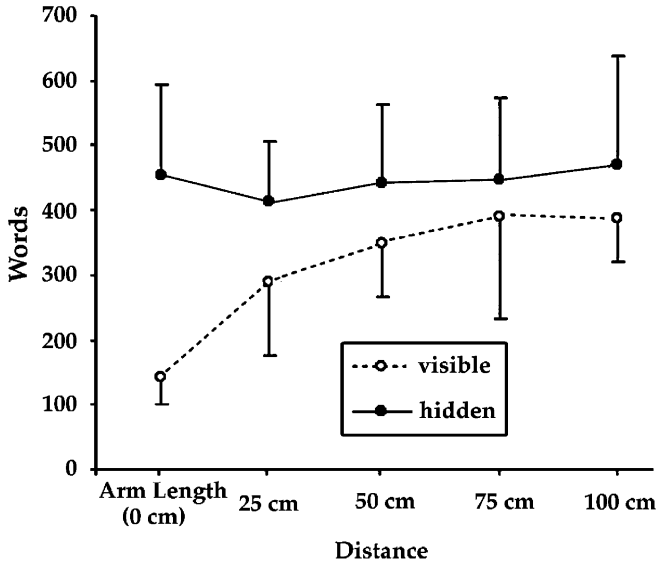

Fig. 3. Mean total number of words used per array as a function of visibility and distance.

due to use of points with deixis: Points with deixis were accompanied by feature descriptions 71% of the time, whereas other points were accompanied by feature descriptions 90% of the time, $\chi^2(1, N = 499) = 23$, $p < .001$.

It has been reported (Butterworth & Itakura, 2000) that pointing is implicated in large-scale attentional shifts. In the current study, pairs identified one target after another, and the distance between successive targets is a measure of the shift of visual attention. Did participants point more often to targets far from their predecessor than to targets closer to their predecessor? The target-predecessor distance was computed for all targets except the first one in each array (M = 19.6 cm, SD = 7 cm).[1] The number of pairs using points with deixis to refer to a target correlated positively with this distance (r = .35, $p < .05$), whereas the number of pairs using other points did not (r = −.1, n.s.), further supporting the third hypothesis.

Difference Between Points With Deixis and Other Points

A fundamental difference emerged between the two kinds of pointing gestures. One explanation for this finding is that demonstratives focus attention, by directing the addressee's gaze to the speaker's gesture (Bühler, 1965). When a gesture carries the main informational burden of a referring act, speakers need to be sure that addressees are attending to it. Using a demonstrative accomplishes this. This explains the reduced verbal effort observed when pairs used points with deixis, and is consistent with the tight coordination between pointing and speech (Marslen-Wilson et al., 1982) and the fact that demonstratives encode joint-attention status (Özyürek & Kita, 2002). It is not clear what function, if any, was served by other points, although they may have been redundant.

[1]Unit of analysis was the target (n = 45). Data were collapsed over distances of the array.

CONCLUSION

Pointing gestures were used to achieve joint attention, by circumscribing referential domains, as well as to identify referents. Claims that pointing is ambiguous, or redundant with language, are often based on the assumption that pointing is exclusively used for identifying referents. In contrast, this study supports the view that language and gesture are used flexibly and opportunistically in dialogue.

Acknowledgments—I thank the Swiss National Science Foundation (8210-061238) and the Office of Naval Research (N000140010660) for support, Mark Steyn for the photographs used in the study, and Zenzi Griffin, Franciska Krings, Teenie Matlock, and especially Herb Clark for comments.

REFERENCES

Argyle, M., & Cook, M. (1976). *Gaze and mutual gaze*. Cambridge, England: Cambridge University Press.

Baron-Cohen, S. (1995). The eye direction detector (EDD) and the shared attention mechanism (SAM): Two cases for evolutionary psychology. In C. Moore & P.J. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 41–60). Hillsdale, NJ: Erlbaum.

Bavelas, J.B., & Chovil, N. (2000). Visible acts of meaning: An integrated message model of language in face-to-face dialogue. *Journal of Language and Social Psychology*, *19*, 163–194.

Beun, R.-J., & Cremers, A.H.M. (1998). Object reference in a shared domain of conversation. *Pragmatics & Cognition*, *6*, 121–152.

Brennan, S.E. (1995). Centering attention in discourse. *Language and Cognitive Processes*, *10*, 137–167.

Brown-Schmidt, S., Campana, E., & Tanenhaus, M.K. (2002). Reference resolution in the wild: Online circumscription of referential domains in a natural, interactive problem-solving task. In W. Gray & C. Schunn (Eds.), *Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society* (pp. 148–153). Mahwah, NJ: Erlbaum.

Buchler, J. (Ed.). (1940). *Philosophical writings of Peirce*. London: Routledge and Kegan Paul.

Bühler, K. (1965). *Sprachtheorie*. Stuttgart, Germany: Fischer.

Butterworth, G., & Itakura, S. (2000). How the eyes, head and hand serve definite reference. *British Journal of Developmental Psychology*, *18*, 25–50.

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

Clark, H.H. (2003). Pointing and placing. In S. Kita (Ed.), *Pointing: Where language, culture and cognition meet* (pp. 243–268). Hillsdale, NJ: Erlbaum.

Clark, H.H., & Marshall, C.R. (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, England: Cambridge University Press.

Clark, H.H., Van Der Wege, M.M., & Katz, A.R. (2002, November). *Pointing in dialogue*. Paper presented at the annual meeting of the Psychonomic Society, Kansas City, MO.

Engle, R.A. (1998). Not channels but composite signals: Speech, gesture, diagrams and object demonstrations are integrated in multimodal explanations. In M.A. Gernsbacher & S.J. Derry (Eds.), *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society* (pp. 321–327). Mahwah, NJ: Erlbaum.

Fleiss, J.L. (1981). *Statistical methods for rates and proportions*. New York: John Wiley.

Grice, H.P. (1975). Logic and conversation. In P. Cole & J.L. Morgan (Eds.), *Syntax and semantics 3: Speech acts* (pp. 41–58). New York: Seminar Press.

Grosz, B., & Sidner, C. (1986). Attentions, intentions and the structure of discourse. *Computational Linguistics*, *12*, 175–204.

Langton, S.R.H., & Bruce, V. (2000). You must see the point: Processing of cues to the direction of social attention. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 747–757.

Levy, E.T., & McNeill, D. (1992). Speech, gesture, and discourse. *Discourse Processes*, *15*, 277–301.

Lyons, J. (1981). *Language, meaning and context*. Glasgow, Scotland: Fontana/Collins.

Marslen-Wilson, W., Levy, E., & Tyler, L.K. (1982). Producing interpretable discourse: The establishment and maintenance of reference. In R.J. Jarvella & W. Klein (Eds.), *Speech, place and action: Studies in deixis and related topics* (pp. 339–378). Chichester, England: John Wiley.

McNeill, D. (1985). So you think gestures are non-verbal? *Psychological Review*, *92*, 350–371.

Olson, D.R. (1970). Language and thought: Aspects of a cognitive theory of semantics. *Psychological Review*, *77*, 257–273.

O'Neill, D.K., & Topolovec, J.C. (2001). Two-year-old children's sensitivity to the referential (in)efficacy of their own pointing gestures. *Journal of Child Language*, *28*, 1–28.

Özyürek, A., & Kita, S. (2002). *Joint attention and distance in the semantics of Turkish and Japanese demonstrative systems*. Unpublished manuscript, Max-Planck Institute for Psycholinguistics, Nijmegen, The Netherlands.

Pechmann, T., & Deutsch, W. (1982). The development of verbal and nonverbal devices for reference. *Journal of Experimental Child Psychology*, *34*, 330–341.

Sacks, H., Schegloff, E.A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking in conversation. *Language*, *50*, 696–735.

Schmauks, D. (1991). *Deixis in der Mensch-Maschine-Interaktion* [Deixis in human-computer interaction]. Tübingen, Germany: Max Niemeyer.

Schmidt, C.L. (1999). Adult understanding of spontaneous attention-directing events: What does gesture contribute? *Ecological Psychology*, *11*, 139–174.

Thompson, L.A., & Massaro, D.W. (1986). Evaluation and integration of speech and pointing gestures during referential understanding. *Journal of Experimental Child Psychology*, *42*, 144–168.

Velichkovsky, B.M. (1995). Communicating attention: Gaze position transfer in cooperative problem solving. *Pragmatics and Cognition*, *3*, 199–224.

(Received 2/26/03; Accepted 6/16/03)