Spatial perspective-taking in conversation*

Michael F. Schober

Department of Psychology, Graduate Faculty, New School for Social Research, 65 Fifth Avenue, New York, NY 10003, USA

Received January 14, 1991, final version accepted November 2, 1992

Abstract


Speakers can describe the locations of objects from their own perspective ("on my left" or "on the left"), their addressee's ("on your right" or "on the right"), or some perspective that avoids choosing one or the other person ("closer to both of us"). This study shows that speakers set spatial perspectives differently with actual conversational partners than with the usually studied imaginary addressees. Speakers with partners tended to use more egocentric perspectives than solo speakers. Pairs varied idiosyncratically in the perspective-setting strategies they picked, but all engaged in the same collaborative process: talking until both were sure they had understood each other. When conversational roles switched, the new speakers allocated spatial perspectives with remarkable precision, taking their partners' perspectives just as often as the partner had taken theirs. Speakers were more explicit about whose perspective they were taking when they held the floor for only one description than when they gave many descriptions in a row.

Introduction

People engaged in conversation often have conflicting perspectives on the topics under discussion. They can differ in basic political or religious assumptions, in

*This research was supported in part by a National Science Foundation Graduate Grant and a Stanford University Dissertation Grant. Many thanks to Herbert Clark, Susan Brennan, Barbara Fredrickson, Monica Schober, Holly Taylor, Barbara Tversky, Elizabeth Wade, and two anonymous reviewers for comments on the manuscript, and to Pam Smul and Heidi Clark for help as research assistants.

0010-0277/93/$06.00 © 1993 – Elsevier Science Publishers B.V. All rights reserved.
their conceptions of what really happened or what an object really is, or in physical vantage points on a scene (see Graumann, 1989). A speaker who is aware of (or wary about) a perspective difference must choose whether to take her\(^1\) addressee’s perspective, to use her own, to use a description true from both points of view, or to use some more “neutral” perspective, if there is one. How does she do this?

The following study focuses on how people describe locations when they have different physical vantage points. This has the advantage that we (as observers) can judge without controversy whose perspective a speaker has chosen at any moment, and we can be sure that speakers always know each other’s vantage point.

**Perspective in spatial descriptions**

Speakers describing locations actually have (at least) two different kinds of perspective choices: spatial and conceptual. Although spatial perspective is the study’s focus, conceptual perspective gives an important contrast.

Spatial perspective refers to the physical point of view reflected in a description of location. The vocabulary describing spatial perspective options varies (see Retz-Schmidt, 1988), but the most frequent distinction is between deictic and intrinsic use. As Miller and Johnson-Laird (1976, p. 396) define it, the deictic system is “the linguistic system for talking about space relative to a speaker’s egocentric origin and coordinate axes” (see also Fillmore, 1977; Levelt, 1989; Lyons, 1977). Speakers can use primary deictic expressions (I, you, here, there, now, then) which refer to basic coordinates of communication (speaker, addressee, place and time of utterance).\(^2\) Or they can use secondary deictic expressions (left, right, front, back), which don’t refer to the basic coordinates but still imply a situated speaker (Ullmer-Ehrlich, 1982).

Deictic use contrasts with intrinsic use, where the point of view is that of any entity other than the speaker that has intrinsic parts (a front, a back, a left, a right) and a canonical orientation. So “behind the chair” is considered an intrinsic description when it deals with the chair’s intrinsic front and back, and the speaker’s location has nothing to do with the description. Descriptions that take the addressee’s perspective, like “on your left”, are also considered to be intrinsic uses.\(^3\) Addressees, like other objects in the world, have intrinsic parts and canonical orientations.

\(^1\)To avoid pronoun confusion, I refer to speakers as female and addressees as male.

\(^2\)To be complete, we should also distinguish between speaker’s and addressee’s place, as well as speaker’s and addressee’s time of utterance.

\(^3\)Some researchers disagree with this scheme, and think of all person-centered descriptions as deictic. For example, Saile (1984) considers “Das Taxi steht, von dir aus gesehen, hinter dem Lieferwagen” (“The taxi is, seen from your point of view, behind the station wagon”) to be a deictic description (cited in Retz-Schmidt, 1988, p. 99).
Spatial perspective-taking

Conceptual perspective (also called “schematization” by Talmy, 1975, 1983; and simply “perspective” by Clark & Schobert, 1991; Clark & Wilkes-Gibbs, 1986; Fussell & Krauss, 1992; Wilkes-Gibbs & Kim, 1991; and others) refers to the perspective inherent in all descriptions. In spatial descriptions, different “locative” expressions (location descriptions) imply particular ways of conceiving of the relevant snippet of the world. For example, a description of an object C as “to the left of” another object D implies an implicit path (getting to C requires a mental movement), while a description of the same object C as “at the left of” D implies only a static relationship. A description of an object E as “the one in front” has a different conceptual perspective than a description of E as “the closest one”; each focuses on different aspects of E’s position relative to other objects.

Speakers’ preferences: Claims and experimental evidence

Which spatial perspective choice do speakers prefer? Most researchers agree with Levelt (1989, p. 52) that the “dominant or default” system for speakers is the deictic system. It is probably easier for speakers to use than the intrinsic system (Miller & Johnson-Laird, 1976, p. 402), and speakers should choose it when in doubt, or without influences to the contrary (Herrmann, 1988).

Speakers should modify their natural egocentrism when their addressees “have difficulties in reconstructing the speaker’s point of view with respect to the scene” (Levelt, 1984, p. 355), since addressee-centered intrinsic uses are probably easier for addressees to understand (Miller & Johnson-Laird, 1976). The nature of the scene makes a difference too: in Miller and Johnson-Laird’s (1976, p. 399) formulation, if a landmark in the scene has intrinsic parsi, addressees tend to interpret a spatial description intrinsically unless the speaker explicitly informs them to the contrary.

Two sorts of laboratory studies have investigated speakers’ descriptions of spatial layouts. In the first sort, subjects typically describe a scene without being told anything about an addressee, or with the implicit supposition that the listener will have the same vantage point. These studies show that speakers prefer using the intrinsic system when describing a relation between two fixed objects, where one object has an intrinsic front (Wunderlich, 1981) or is moving (Levelt, 1984). Speakers describing their apartments and the rooms in them give “tours” from the perspective of a visitor traveling within the scene or from where the visitor’s gaze would travel (Linde, 1979; Linde & Labov, 1975; Ullmer-Ehrich, 1982). Speakers verbally negotiating a path through a “spatial network” made of nodes and lines use more deictic than intrinsic expressions (Levelt, 1982). In these studies, speakers tend to use the same perspectives consistently (Ehrich & Koster, 1983; Levelt, 1982; but see Taylor & Tversky, 1990, for a contrary finding).

In the second sort of study, carried out by Herrmann and his colleagues (e.g.,

Subjects viewing CRT screen displays are told about the vantage point of an imaginary partner. Each display contains a target object, another (reference) object, an indicator of the subject's own vantage point (always at the bottom of the screen) and an indicator of the vantage point of a hypothetical addressee (usually a different position on each display). The subjects describe the location of the target object relative to the reference object, using short phrases like “rechts vom Tisch” (“to the right of the table”). (Either the subjects are instructed to limit themselves to what Herrmann considers the four basic localizing prepositions – “in front of”, “behind”, “left of”, and “right of” – or Herrmann only selects these four for analysis.)

In one study (Bürkle et al., 1986), speakers used few deictic expressions when their vantage points differed from their addressees’. In another study (Gräf, described in Herrmann, 1988), students took their addressee’s perspective less than 50% of the time when they described a plant’s position in a room for a puppet that was supposed to be a fellow student. This proportion increased to 90% for a child puppet, and about 75% for a professor puppet. To explain this, Herrmann suggests that speakers take their addressees’ perspectives in order to reduce the burden of perspective-taking for them.

In another CRT study (Herrmann et al., 1987), students were instructed to take the perspective of their imaginary addressee as they described the location of a “1” on the screen. Speakers were quickest when they shared the vantage point with their addressee, slower at 90° and 270° offsets, and slowest when the addressee was facing them (180° offset). The authors concluded that the cognitive effort for addressee-centered intrinsic use depends on where, relative to the speaker, the listener and the target object are.

**Perspective in conversation**

In all these studies, speakers give monologues for imaginary addressees. But conversation is different from monologue. People engaged in conversation tend to say less than in monologues (Krauss & Weinheimer, 1966; Levelt, 1989). In conversation, references are complete only when they are grounded, or when both participants in the conversation are satisfied that what has been said has been understood (Clark & Schaefer, 1987, 1989; Clark & Wilkes-Gibbs, 1986; Schober & Clark, 1989; Wilkes-Gibbs, 1986). Speakers and addressees use

4Although a number of developmental studies of children’s egocentrism also distinguish between people’s physical vantage points (e.g., Cox, 1977; Fehr, McMahon, & Fehr, 1982; Flavell, 1974; Flavell, Botkin, Fry, Wright, & Jarvis, 1968; Foursman, Leiber, & Fernie, 1984; Jacobsen & Waters, 1985; Piaget, 1959; Schachter & Gollin, 1979), they do not focus on linguistic descriptions.
Spatial perspective-taking

various strategies to exchange evidence of understanding for each contribution to a conversation. For example, they nod, look puzzled, or ask for more information, and they interrupt the flow of conversation to repair their own and their partners' errors and misunderstandings.

Conversational partners try to minimize their collaborative effort—the work they both have to do in order for the reference to succeed (Clark & Wilkes-Gibbs, 1986; Wilkes-Gibbs, 1986). This is very different from what traditional theories of least effort predict, where a speaker picks the shortest noun phrase that will enable her addressee to pick out the referent in context (see Clark & Wilkes-Gibbs, 1986). Instead, the speaker designs her utterance in a way that will ultimately get the reference understood most efficiently. For example, she might pick a longer noun phrase to describe an object than a simple notion of least effort would predict, because she believes that by doing so she won't have to refashion her description later so her addressee can get it. She might present a complex noun phrase in installments, making sure each part is understood. Or she might try out a description and allow her addressee to refashion if it doesn't work (Clark & Wilkes-Gibbs, 1986; Wilkes-Gibbs, 1986).

Speakers of monologues don't have these same resources available. In a sense, a speaker in conversation can get away with more than a speaker giving a monologue, because she can rely on her addressee to point out any lapses in clarity. In another sense, a speaker in conversation must perform to a higher standard, because if her addressee is willing to put in the effort to understand, no lapses in clarity will be tolerated.

One could argue that collaborative processes simply ease social relationships, make conversation smooth, or serve to motivate the conversational participants. But a recent study (Schober & Clark, 1989) shows that collaboration actually influences a listener's understanding. In the study, addressers who could give collaborative feedback to their partners matched ambiguous figures correctly virtually 100% of the time, while overhears who heard all the same words matched fewer figures correctly. This was true even though the subjects were all strangers, the figures were novel, and the overhears heard every word the conversational partners uttered from the moment they met each other. Even when overhears could have as much time as they wanted to listen to the conversation, pausing a tape-recording when they wished, their task performance was inferior to the addressers'.

How do conversational partners set spatial perspectives? One possibility is that they follow the same principles as they do in coordinating on conceptual perspectives. For conceptual perspectives, conversational partners tend not to discuss their choices explicitly, but rather they tend to be locally consistent in the ways they refer to objects, acting according to “input-output coordination” (Garrod & Anderson, 1987). If one partner offers a particular way of conceiving of an object as “input”, the other uses the same conceptual perspective as
“output”. Garrod & Anderson claim that, in general, once coordination has been achieved “no modification can be introduced without violating the principle of local consistency” (p. 209).

To study spatial perspective in conversation we must modify the usual deictic/intrinsic classification scheme. Take the following interchange:

A: It's on your left.
B: On my left, okay.

In the usual scheme, A’s utterance is addressee-centered intrinsic, and B’s is deictic; we treat each utterance in isolation and classify it according to the viewpoint of the speaker. But on some level, the utterances are both from the same perspective (B’s). To account for this, I treat the pair of conversational partners as the unit of analysis, and classify each utterance according to which partner’s perspective is taken. Thus, utterances by either partner can be from A’s perspective, from B’s, from an ambiguous perspective (true for both), or from a neutral perspective (one that would be true no matter where the partners were located).

The study

This study examines a simple conversational situation: speakers point out the location of one object relative to another. The task and the displays are similar to those Herrmann and his colleagues have used, but there are two important differences. First, subjects are not constrained to use any particular localizing phrases in any particular sentence frame, and so they can use locative descriptions beyond the four (left, right, front, behind) that Herrmann studied. Second, the subjects must describe one of two identical objects. I suspect that unconstrained speakers describing dissimilar objects would name the objects rather than describing their locations.

The objects to describe in this study are two circles. So subjects have three perspective choices: egocentric (deictic), addressee-centered (intrinsic), or an ambiguous perspective true for both (deictic and addressee-centered intrinsic). Object-centered intrinsic descriptions are not an option, as the circles have no intrinsic perspectives.

Consider A’s ambiguous description “to the right” under the following conditions: A and B are both facing two objects, one closer to A and one closer to B. B is positioned 90° to A’s right. The description is true from A’s (deictic) perspective, but it is also true from B’s (addressee-centered intrinsic) perspective. B doesn’t even need to know which perspective A intended, as either one works.
The study addresses the following questions:

(1) **How are the perspectives speakers set with actual addressees different from those speakers set for imaginary addressees?** If interaction makes a difference, then an addressee’s feedback should affect the perspectives speakers choose. Speakers with partners should adapt to what works best for both of them, and those with imaginary addressees should be more uniform in sticking with the perspectives they start with.

(2) **How does the difficulty of taking the other’s perspective affect the perspectives the speaker chooses?** As Herrmann’s results and studies of children suggest, it is trivial for A to step into B’s shoes (figuratively) if they share the same vantage point, but it is more difficult if they are farther offset. Both Levelt (1989) and Miller and Johnson-Laird (1976) predict that at greater offsets egocentric use will dominate (especially when object-centered intrinsic perspectives are not among the options).

But egocentric use poses a collaborative problem. A knows that B faces perspective-taking difficulties similar to hers; at greater offsets, A should use the means available to reduce not only her own effort, but the collaborative effort of perspective-taking. I predict that at greater offsets A should use both-centered perspectives whenever possible, or else substitute descriptions whose conceptual perspective is less problematic.

(3) **Whose perspective is taken when conversational roles switch?** If A has taken B’s point of view exclusively all along, whose perspective will B take when he describes locations to A? If B is egalitarian (follows a principle of reciprocity), he will return the favor and take A’s perspective. Alternatively, if B follows Garrod and Anderson’s (1987) input–output coordination principle, he should continue in the perspective A has been taking thus far.

(4) **How does the length of time a speaker has the floor affect the perspectives she chooses?** How B resolves the speaker change problem may depend on how long he will be speaking (how much he has to describe). If he only has the floor briefly it may not be worth his while to renegotiate perspectives. But if he intends to talk longer, A may be happier if he expends the effort of taking her perspective.

To answer Question 1, I compare the perspectives chosen by people who have an imaginary addressee and those who have an actual addressee who gives feedback. To answer Question 2, I compare the perspectives chosen by speakers whose vantage points are closer and farther from their addressees’ vantage points. To answer Question 3, I compare the perspectives that the two speakers in a pair choose. And to answer Question 4, I compare the perspectives chosen by speakers who describe many locations one after another to the perspectives chosen by speakers who alternate describing a location and having a location described to them.
Method

Forty students performed a modified version of a referential communication task (Krauss & Weinheimer, 1964, 1966). The students were native speakers of American English who received either course credit or payment. None of the pairs of students knew each other beforehand.

In the task, one student, the director, described which of two circles in a display had an X on it for another student, the matcher, who was either imaginary or actually present. The matcher's task was (or would have been, for the imaginary matcher) to mark with an X the circle on his display that corresponded to the director's marked circle. Each director (and matcher, if there was one) went through a series of 32 of these descriptions.

Each of the 32 different displays had two circles within a round frame, located next to each other. On each display the circles were positioned in one of four ways: (1) vertically, (2) horizontally, (3) diagonally (bottom left and top right), and (4) diagonally (top left and bottom right). At the bottom of each page (the side closest to the viewer) was an arrow and the word “You”, showing the viewer’s vantage point. Another arrow, along with the word “Partner”, showed the other person’s viewpoint. The viewer’s and partner’s vantage points were either (1) the same, (2) off by 90°, (3) off by 180°, or (4) off by 270° (see Figure 1 for samples).

Subjects saw one of eight random sequences of these 32 displays. Four were the reverse of the other four, so that displays early in one sequence would be late in another. Each student played the game in one of three ways:

1. Solo (eight students): the student, alone in the room, described each of the 32 displays into a tape-recorder for an imaginary addressee. This condition is similar to what Herrmann’s subjects experienced.
2. Interactive with long turns (eight pairs of students): the students could speak with each other. One student was the director for one sequence of 32 displays, and then the students switched roles: the other student was the director for another sequence of 32 displays.
3. Interactive with short turns (eight pairs of students): the students could speak with each other. Each described a different sequence of 32 displays, but the displays were interleaved so that the students had to switch roles after every display.

The experiment room was set up the same for all students, except that the pairs (conditions 2 and 3) sat on either side of a visual barrier, so that their actions would have to be confined to verbal ones. The students sat at a table, and the

6To continue avoiding pronoun confusion, I treat directors as female and matchers as male.
Spatial perspective-taking

Figure 1. Sample displays.
The experimenters instructed the students about the nature of the task, and told them they were to say anything they needed to get the task done. The students were also told not to move the notebooks on the table—the arrow saying “You” was always supposed to be closest to them (from observing videotapes later, I caught the pair that cheated and discarded their data). The solo speakers were told to describe each location for an imaginary partner who had to draw the X on the correct circle, but they weren’t told anything specific about what their imaginary addressees were like.

Setting the experiment up this way had its disadvantages. Since students in pairs were seated facing (though hidden from) each other, many times their relative positions on the displays were different from their actual physical positions. No students ever mentioned any discomfort with the procedure, and I believe this possible disadvantage is outweighed by the set-up’s advantages. While conversation was unrestricted, all communication had to be verbal, and so transcripts of conversation and videotape of hand movements gave a complete account of students’ actions. The experiment could be videotaped unobtrusively because students remained seated in one place throughout the experiment; this would have been unwieldy (if not impossible) if students had to be repositioned 32 times, and the students would likely have found the experience awkward and tiresome.

The students were both audio taped and videotaped. A low level of white noise played in the room throughout, so that subjects couldn’t take the sound of matchers’ pen scratches as evidence of understanding. After the students finished, they answered on a questionnaire whether or not they knew their partner, and how easy or hard they had found the task and why.

### Coding

From transcripts of the speakers’ monologues, I extracted the locative expressions and categorized them according to spatial and conceptual perspective (see Appendix for examples of each). Spatial perspectives were either director-centered, matcher-centered, or “both-centered” (true from both people’s points of view). A very small percentage didn’t fit into any of these categories.

The two main types of conceptual perspective were **static** and **implicit path** locatives (these distinctions are partly inspired by Talmy, 1983). Static locatives don’t imply a path of motion, as in “on the right” or “the lower left circle” or “at

---

7In piloting the study, I found that even the most silent pens made a scratching noise as the matchers wrote the X’s on their sheets, and pairs of subjects resorted to using these sounds in place of verbal feedback. While this is interesting in its own right, addressees normally give verbal feedback in the perspective-setting cases I’ve set out to study.
the left of you”. Path locatives do: “to the right of you”, “to the lower left” or “closer to you”. There were also a few directional phrases, with a very explicit path of motion (“the one you get to first”), and a very few that didn’t fit into the other categories (“the same one as last time”).

Results and discussion

Question 1: Speakers with partners versus solo speakers

As others have found (e.g., Krauss & Weinheimer, 1966;Levelt, 1989), speakers with partners said substantially less than the solo speakers, 10.9 words/display versus 18.2 words/display), $F(1, 22) = 15.47, p < .0001$. Each speaker in a pair also used fewer locative expressions than the solo speakers, averaging 1.28 during her respective sequence, compared with the solo speakers’ 2.11, $F(1, 22) = 17.60, p < .0001$. Only one of the speakers in the 16 pairs ever talked about the location of the non-target circle, but three of the eight solo speakers did this consistently.

Both solo speakers and speakers in pairs were highly accurate in their descriptions; only rarely did any speaker use a locative whose perspective didn’t correspond to the one the speaker claimed she was using, or that didn’t make sense from any possible perspective. We can’t conclude much from the six “errors” made by solo speakers and the ten made by speakers with partners, except to note that 14 of these 16 errors happened when directors’ and matchers’ vantage points were offset.

Matchers were remarkably accurate (one error each by two students from among the 16 pairs, both for a diagonally oriented display at a 90° offset). Directors’ errors didn’t necessarily lead to trouble for matchers. Directors would usually catch their mistakes, or the matchers would point out the mistakes and ask for more information. But in a few cases, matchers correctly marked their displays based on impossible descriptions, without voicing any objections. They acted as if opening up negotiation about perspective would be more trouble than it was worth.

8This is true no matter how we compare them— as individuals, as a pair, or as an average of the individuals in the pair. As director, the first speaker in each pair used 1.30 locatives per display, which is reliably less than the solo speakers, $F(1, 22) = 17.68, p < .0001$. The same is true for the second director in each pair, 1.26 to 2.11, $F(1, 22) = 15.89, p < .0001$. If we reason that we should include all the locatives said by both director and matcher in the first sequence, because the matchers contributed too, we get the same difference, 1.39 to 2.11, $F(1, 22) = 13.075, p < .002$. And the same is true for the second sequence, 1.40 to 2.11, $F(1, 22) = 8.73, p < .007$. As a general rule, when I report a pair’s scores, I will use the average of each of the directors’ scores, unless I specify otherwise.
Conceptual perspective

All the speakers used more than just the four basic localizing expressions that Herrmann studied. Only 52% of the solo speakers’ locatives and only 61% of the speakers in pairs contained the words “left”, “right”, “front” or “behind” (this includes odd utterances like “to the rightmost of you”), ranging from 30% to 77%. Some speakers used descriptions like “above”, “below”, “upper”, “lower” and “back”. But most of the other localizations were of the closer-farther sort (“the farthest one”, “further away from you”, “the closest circle” and “nearest”).

Speakers didn’t always use consistent conceptual perspectives. For example, in one description a student located the non-target circle as “directly in front of you”. Later, when that same circle was the target circle, he described it as “the first one” or “the one below”. Speakers didn’t even limit themselves to one conceptual perspective at a time, as we see in “farther away, just a little bit behind the first small circle” (Taylor & Tversky, 1990, report similar findings for solo speakers describing map layouts).

Spatial perspective

Speakers with partners set spatial perspectives differently from solo speakers (see Table 1). They used more egocentric locatives, 13% versus the solo speakers’ 1.7%, $z = 7.53, p < .0001$, Wilcoxon rank sum test. Correspondingly, they used fewer matcher-centered locatives, 66% versus 77%, $z = 4.84, p < .0001$, Wilcoxon rank sum test.

These percentage differences underestimate the difference between solo speakers and speakers with partners. Solo speakers were almost uniform in being non-egocentric; none used more than 6.1% egocentric locative expressions, and several never used their own perspective at all. In contrast, the speakers with partners varied widely; almost all used at least a few egocentric locatives, up to 100% of the time (96%, if we average the percentages for both speakers in a pair). The difference in variability is clear if we compare the standard deviations of the arcsin-transformed proportions, .07 to .26, $F(7, 15) = 13.27, p < .005$, using the standard equality of variances test. This is good evidence that setting perspectives interactively is different from setting them for a tape-recorder.

Solo speakers who used egocentric perspectives often apologized. “Place the X

Table 1. Perspectives taken overall

<table>
<thead>
<tr>
<th></th>
<th>Matcher’s</th>
<th>Both</th>
<th>Director’s</th>
<th>Other</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solo speakers</td>
<td>77</td>
<td>21</td>
<td>2</td>
<td>1</td>
<td>100</td>
</tr>
<tr>
<td>Pairs (directors)</td>
<td>66</td>
<td>20</td>
<td>13</td>
<td>1</td>
<td>100</td>
</tr>
</tbody>
</table>
in the upper left hand. wait wait wait, sorry about that, delete that. Place the X in the upper right hand circle.”9 They acted as if their egocentric uses were mistakes or left-right confusions, not intentionally egocentric descriptions. In contrast, speakers with partners were not apologetic; some speakers spoke egocentrically even when they shared the addressee’s vantage point: “It’s on my left” and “From my point of view it’s the one on the bottom.”

Speakers, with or without partners, most often took their addressee’s perspective. If we add in the both-centered perspectives, this means that the overwhelming majority of locative phrases were true from the addressee’s point of view. In fact, speakers used addressee-centered locatives even when they shared their addressee’s vantage point, for example, using “it’s the one on your left” even though “my left”, “our left” and “the left” were equally true. This contrasts with other findings that speakers prefer egocentric localizations, especially when there’s no reason to avoid them. Perhaps the particular display and set of perspective options in this study accounts for this; in any case, it is clear that the often-observed egocentric bias in speakers is not absolute.

The important point here is that speakers with partners are not uniform in taking their addressee’s perspectives, while the solo speakers are. Why? Is it simply the presence of a live partner that makes the difference, or does the process of interacting conversationally affect perspectives?

**How pairs coordinate perspectives**

Most pairs never discussed explicitly whose spatial perspective they should use, just as Garrod and Anderson (1987) found with conceptual perspectives. With only two exceptions, speakers plunged right in offering descriptions from one particular perspective. Most often, matchers accepted these descriptions without complaint, either immediately, as in

D. Uh, it’s the one on the right.
M. Gotcha.

D. And this one, is the bottom left hand circle for you?
M. Okay.

or by repeating the locative to ascertain that the offered perspective was understood.

---

9The following conventions represent speech in this paper: a period with spaces on either side (.) indicates a pause. Asterisks enclose bits of overlapping speech. A colon indicates a lengthening of a sound. A dash following a word (“the circ-”) indicates that the word was cut off in midstream or not quite completed.
If all the interchanges were this simple, we could conclude that the interactive process doesn’t change speakers’ perspectives. But they were not. In 26 descriptions (out of the 1024 in the corpus of the 16 pairs’ conversations), the matcher took a more active role. These cases are of particular interest, because if the matcher’s intervention matters, we should see the effects on subsequent descriptions.

These 26 cases break down into five categories, presented here in order of frequency and with an example of each:

(1) Matcher accepts director’s description conditionally, and reframes it in a new perspective as a check (12 occurrences).

D. And it’s farthest from me?
M. Closest to me.
D. Mm-hm?

(2) Matcher accepts conditionally, and offers more information in the same perspective as a check (6 occurrences).

D. It’s the one in front.
M. Closer to me.
D. Right.
M. Okay.

(3) Matcher questions whose perspective director means (6 occurrences).

D. Okay this one I ha- I have- there are . equal horizontal?
M. *Mm-hm*.
D. *And I* have one in the right hand. *So-.*
M. *In* in your right hand or .
D. In your right.
M. Oh, okay. You’re making this really easy for me.

(4) Matcher rejects by reframing in a different perspective (1 occurrence).

D. The: second page . um . from . your point of view, it’s um . n- uh I’ll do it from my point. From my point of view it’s in the um . circle . to the left or the west.
M. Left, okay. Isn’t that the s- do we have the same point of view *on that?*
D. *Yeah*, *we have* the same point of view.
M. *Okay*.
D. So *it’s the* same way.
Spatial perspective-taking

M. *Okay,* so if it’s in the left, that’s fine.
D. Okay.

(5) The matcher elicits a repair by not accepting (1 occurrence).
D. Uh this one’s your top left?
M. U:h.
D. I’m sorry your bottom. bottom left.
M. Okay. I don’t have a top left (laughs).
D. (laughs).

Although the sample of these negotiations is small, the direction of a pair’s future descriptions should be predictable. I propose that the perspective that “sticks” – that the director now adopts – will be the one that both partners have agreed upon.

In the 12 examples in case (1) the director and matcher agreed to two different perspectives. Directors stuck with their original perspectives on 72% of these subsequent descriptions. If we include the subsequent both-centered descriptions in with either matcher or director-centered descriptions, the figure rises to 89%.

In the six instances (from four pairs) in case (2), both partners accepted one perspective. Directors stuck with the perspectives they had negotiated 97% of the time. Only one director ever uttered a locative from a different perspective than one already negotiated, and it was only as an addition to the one from the negotiated perspective she’d already uttered. The matchers accepted each of these subsequent locatives without question. Case (4) is similar to case (2), but what was accepted was not what the director originally presented. Every single one of the director’s eight subsequent localizations at that vantage point offset were both-centered.

So, to speculate based on only a few cases, it seems that speakers prefer to stick to the perspective they’ve already used if their addressees accept them, and only change perspectives when the addressee rejects their offerings outright.

**Question 2: Perspective at different offsets**

As Table 2 shows, neither solo speakers nor speakers with partners used different perspectives at different degrees of offset. They took their partner’s perspective or a both-centered perspective almost every time. Speakers in pairs did use slightly more egocentric locatives at greater offsets: 8% at the shared vantage point, 14% at 90° and 270° offsets, and 16% at the 180° offset, $F(2, 30) = 4.79$, $p < .016$, using arcsin-transformed proportions. But using the Tukey method of multiple comparisons, we see that the only reliable difference is between the 8% at the shared vantage point and the 16% at the 190° offset, $q = 4.46$, $p < .01$; this
These few cases are when the speaker used a both-centered locative phrase to describe the orientation of both circles, as in “the circles are diagonal, top left and bottom right”.

difference is unsurprising, since speakers used mostly both-centered descriptions at the shared vantage point.

What accounts for this failure to replicate Bürkle et al.’s (1986) and Herrmann’s (1988) results? Perhaps the lack of an object-centered intrinsic perspective option altered speakers’ preferences. Or perhaps the task was so easy that no offset was quite hard enough to require perspective change. If this is true, we might still see differences in reaction time at different offsets, as Herrmann et al. (1987) found.

Although spatial perspectives didn’t differ at the different offsets, conceptual perspectives did. As Table 3 shows, solo speakers used descriptions like “closer” and “farther” more often at greater offsets: 28% of the time at the shared vantage point, 29% at 90° and 270° offsets, and 34% at the 180° offset, linear trend $F(1, 21) = 4.35, p < .05$. The same pattern held for the speakers in pairs: 23% at the shared vantage point, 27% at the 90° and 270° offsets, and 29% at the 180° offset, linear trend $F(1, 30) = 7.76, p < .01$.

Table 3. Closer–farther type descriptions

<table>
<thead>
<tr>
<th></th>
<th>Percentage of closer–farther locatives at each offset</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0°</td>
</tr>
<tr>
<td>Solo speakers</td>
<td></td>
</tr>
<tr>
<td>Matcher-oriented</td>
<td>28</td>
</tr>
<tr>
<td>Both-oriented</td>
<td>23</td>
</tr>
<tr>
<td>Director-oriented</td>
<td>23</td>
</tr>
<tr>
<td>Other</td>
<td>23</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
</tr>
</tbody>
</table>

Note: Each pair’s score is the average of each score of each director in the pair.

* It isn’t possible for speakers to localize a circle using a both-centered perspective.
The use of closer–farther descriptions is a sensible solution to greater perspective-taking difficulty at greater offsets. Consider a display with two circles in a diagonal orientation, where for the speaker the target circle is to the left and in front of the other circle. If the addressee is offset 180°, the speaker taking her partner’s perspective would need to describe the target circle as “to the right and behind”, or “to the right and on top”. Technically, the speaker would be correct in naming the target simply “to the right” or “behind”, but this would violate principles of conversational cooperativeness and simplicity (Grice, 1975); saying “in front” or “behind” tends to presume a straight-line front-back relation. “Closer” and “farther” do more with less work, and the speaker doesn’t have to consider her partner’s perspective as carefully. Also, “front” and “back” can be ambiguous, as they can refer to objects’ intrinsic fronts and backs, and can refer to the closer or farther end of an object.

Question 3: Perspective and speaker change

Conceptual perspective

Just as Garrod and Anderson (1987) found, pairs set up their own idiosyncratic conventions for descriptions from the very first utterance, rather than explicitly negotiating conceptual perspectives. One partner would use one sort of description and the other would follow suit, developing a locally consistent, mutually satisfactory description scheme. The proportions of static locatives used by the first speaker in a pair correlated very highly with the proportion used by the second speaker, \( r = .94 \), \( F(1, 14) = 114.05, p < .0001 \). Closer–farther proportions were also nearly identical, \( r = .87 \), \( F(1, 14) = 41.70, p < .0001 \). Speakers in pairs often used exactly the same words to describe the same display, or nearly identical words. For example, one speaker described a display as “on the right bottom”, and her partner later described the same display as “on the bottom right”.

The same high correspondence can be seen within each pair. The proportions of static locatives, path locatives and other locatives used by the first speaker in a pair correlated highly with the same proportions for the second speaker. Correlations for all 16 pairs were positive, ranging from .68 to 1.00, with an average of .96, median .997. The same was true for each speaker’s proportions of closer–farther, front–behind, and right–left descriptions. These ranged from .75 to .9999, averaging .94, median .992.

Spatial perspective

If Garrod and Anderson’s (1987) principle for conceptual perspective applies to spatial perspective, then the second speaker in a pair should use the same
spatial perspective as the first in describing an object. But exactly the opposite happened: if the first speaker described a particular object egocentrically, then the partner described that object egocentrically too. The proportion of director-centered locatives by the first speaker correlated very highly with the proportion used by the partner, $r = .94$, $F(1, 14) = 105.80$, $p < .0001$. Looking at all three proportions within pairs, the first speaker’s proportions in each perspective (director-centered, both-centered, matcher-centered) correlated .98, on average, with the partner’s, ranging from .88 to 1.00.

Speakers thus seemed to follow a norm of reciprocity or fairness, matching their partner’s spatial perspectives (in the opposite direction) with impressive precision.

**Question 4: Long-term versus short-term speakers**

As Table 4 shows, short-term speakers didn’t take reliably different spatial perspectives than long-term speakers. They used 17% director-centered locatives, compared to the long-term speakers’ 8.9%, $F(1, 14) = 0.12$, n.s., using arcsin-transformed proportions. Both groups used the same percentage of matcher-centered locatives (66%) and roughly the same percentage of both-centered locatives, 16% and 24%.

Short-term speakers didn’t match their partners’ spatial perspectives more closely either. The average correlation of the proportion of each kind of locative (director-centered, both-centered, matcher-centered) for speakers within each short-term pair was $r = .995$ (ranging from .985 to 1.00), not reliably higher than the long-term speakers’ average correlation of $r = .96$ (ranging from .88 to .996), $t(7) = 1.93$, n.s.

So the hypothesis that short-term speakers, having greater local pressure, would use different perspectives and match their partners more closely than long-term speakers were not borne out. But short-term speakers were much more likely to mention explicitly whose point of view was being used, for example, saying “From your point of view it’s the one on the left” rather than “It’s the one on the left.” Short-term speakers explicitly mentioned perspective at least once during 48% of the descriptions, compared to 8.4% for the long-term speakers.

<table>
<thead>
<tr>
<th>Perspective</th>
<th>Matcher’s</th>
<th>Both</th>
<th>Director’s</th>
<th>Other</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short-term speakers</td>
<td>66</td>
<td>16</td>
<td>17</td>
<td>1</td>
<td>100</td>
</tr>
<tr>
<td>Long-term speakers</td>
<td>66</td>
<td>24</td>
<td>9</td>
<td>1</td>
<td>100</td>
</tr>
</tbody>
</table>
Spatial perspective-taking

(and 17% for the solo speakers), $F(2, 21) = 4.22, p < .03$, using arcsin-transformed proportions. Short-term speakers also varied much more in explicitness, with a standard deviation of .51 compared to the long-term speakers’ .18, $F(7, 7) = 8.49, p < .01$, using the standard test for equality of variances. This suggests a modification to the finding from both “monologue” and dialogue studies (Garrod & Anderson, 1987; Levelt, 1989, p. 51) that speakers are rarely explicit about perspective: speakers are more explicit when speaker change is more rapid.

Subjective commentary

On the questionnaire at the end of the study, students wrote that having offset vantage points created a dilemma, easily surmounted. “She was very helpful at describing the marks from my ‘perspective’ and so all I had to do was mark ‘X’. When it was time to tell her where to place the mark I had to see from her perspective, but it was not too difficult.” “It seemed to be just a matter of stating where the X was, mainly from my partner’s perspective. (I thought that would be easier for her.)”

Three of the eight solo speakers noted that doing this task without an actual partner was strange (even though they didn’t know they were being compared to pairs). One said “It wasn’t difficult at all, just very awkward talking to someone who’s not there.” One pointed out that she couldn’t be sure of the success of her descriptions: “Relatively easy. Just describe from my partner’s perspective always and it works out. (I hope).” And one was even more explicit: “If I had a partner, we would both have to agree on one method of describing orientation.”

One speaker with a partner, on the other hand, noted the importance of being able to ground her utterances with an actual addressee. “It was pretty easy, although I’m not especially spatially oriented, because my partner and [I] confirmed each other’s directions. This way it seemed hard to get confused or have a misunderstanding.”

Speakers with partners noted that both had adapted to each other. One pair “quickly adapted the pt-of-views that were similar, i.e., the partner’s pt-of-view.” Another found that the task flowed quickly once they had “a uniform rule”. One matcher found the task “easy, once I figured out to change my point of view to that of my partner’s every time she was describing where to put the X”. Another told of how he molded the conversation: “After we started, I eventually realized it would be much easier to define the position in terms of the other person, so I started to push my partner toward that. Once she accepted, we cruised through only using our personal reference when necessary for clarification under confusion. After that, it was very easy.”
Summary

Speakers describing locations on a simple display set spatial perspectives differently with actual addressees than imaginary ones. Some took exclusively egocentric perspectives, while none of the solo speakers did. Pairs varied idiosyncratically in the perspective-setting strategies they picked. Speakers tended to keep using perspectives that they and their partner both had agreed upon—until the roles switched; then the new speakers took their partners’ perspectives just as often as the partner had taken theirs.

The degree of offset between partners didn’t much affect whose spatial perspective was chosen, although it did affect conceptual perspectives. This may well be because the displays in the current study were much simpler than in other studies; indeed, in a study like the current one using a more complex display (Schober, 1990, Study 2), speakers were more egocentric overall when the degree of offset was greater.

The results in this study suggest certain additions to a theory of spatial perspective. First, speakers don’t necessarily use egocentric descriptions as defaults; the predominance of addressee-centered description among solo speakers and some of the pairs in this study argues against this claim. It remains to be seen whether a general-purpose default perspective really exists, or whether features of spatial layouts or of social situations simply allow particular perspectives (egocentric ones) to predominate for most speakers on most occasions.

Second, spatial perspective-taking in conversation is a complicated affair. Speakers do more than take into account their own egocentric bias, their addressee’s relative position, and the nature of the objects they are describing. They also take into account what their addressee gives as evidence of having understood, and how their addressees have already described locations. Establishing understanding in conversation includes the “dynamic and reciprocal interplay of perspective setting and perspective taking” (Graumann & Sommer, 1988, p. 200).

I have attributed the results of this study to high-level pragmatic processes: (1) speakers design their utterances for the particular needs of their addressees; and (2) both partners actively collaborate with each other to ensure understanding. Could one explain the results with more elementary cognitive processes? One could argue, for example, that the partners’ precision in matching each other’s spatial perspectives results simply from lexical priming. Although certainly collaborative processes rest upon elementary cognitive processes, the results of a recent experiment dealing with conceptual perspectives (Wilkes-Gibbs & Kim, 1991) suggest that a higher-level pragmatic account is necessary. In the study, the experimenter gave speakers verbal labels for ambiguous figures. In subsequent conversation, speakers didn’t necessarily use those “primed” labels, but rather worked out descriptions with their particular conversational partners. If only
Spatial perspective-taking

lexical priming determined the choice of expressions in conversation, one would expect speakers to use the descriptors they had already heard.

Many factors I have not investigated here no doubt affect spatial perspective use: speakers' judgements of their partner's knowledge and abilities, assessments of what kinds of perspective shifts are difficult for themselves and their partners, and how polite they want to be. (According to Brown & Levinson, 1987, taking one's addressee's perspective is usually more polite than using one's own, as it imposes less upon the addressee.) To understand such factors, we need to know more about the mental processes involved in putting oneself in another person's shoes. It may be a matter of performing a mental rotation – either of the display or of oneself – analogous to a physical rotation. If this is true, then rotations of 180° should always be harder to achieve than rotation of 90°, etc. Recent results (e.g., Bryant, Tversky, & Franklin, 1992; Herrmann, 1987) suggest that this view may be too simple.

To further complicate the matter, different people have different strategies and abilities for performing mental rotations (cf. Just & Carpenter, 1985), and this may influence their judgements of the difficulty of a particular transformation. People’s general level of egocentrism might change under greater cognitive load (Flavell, 1977; Looft, 1972). And if Davis (1983) is right, perspective-taking ability is a stable individual difference. This might affect whether speakers or addressees are savvy or wary enough to recognize when their perspectives do not coincide. Considerations like these suggest that there is more to learn about spatial perspective-taking.

References


Spatial perspective-taking


Appendix: Coding system

The conversations included some cases in which it is unclear which speaker should be held responsible for a particular utterance, as when one person started an utterance and the other completed it. Although this was rare, some analyses considered all the locatives uttered by a pair, so that decisions about responsibility could be avoided.

**Spatial perspective**

Each locative expression was either director-centered (“d”), matcher-centered (“m”), or both-centered (“b”). “On the left”, for example, would be classified as director-centered if it were only true for the director, matcher-centered if it were only true for the matcher, or both-centered if it were true for both of them. These coding rules were used even if it seemed obvious from context that the speaker meant to be speaking from a different perspective. Explicit mention of one or the other person’s vantage point (“on the left for you” or “on my left”) overrode other coding considerations. In a very few cases the locative was so strange that it couldn’t be classified for perspective; these were not included in the perspective counts.
Conceptual perspective

Each locative was either a static locative, an implicit path locative, a directional phrase, or in a few cases, something else. The following locative expressions are examples of static locatives: in front of us, behind, on the left, right, on our right, in uh your top right corner, on your right hand, and the lower left hand circle. The following are examples of implicit path locatives: closer to me, to the right, farther away, furthest away from me, closest to: both of us, far away, the closest one, to your left, to the up, the furthest to the left, up above the other circle, off to the l-right, and away from you toward your right. The following are examples of directional phrases: the rightmost direction, the first one you come to when you're looking at, and that you will run into first if you move in a straight line.