From: AAAI Technical Report SS-96-03. Compilation copyright © 1996, AAAI (www.aaai.org). All rights reserved.

Addressee- and Object-centered Frames of Reference in Spatial Descriptions

Michael F. Schober

New School for Social Research
Department of Psychology, AL-340
Graduate Faculty
65 Fifth Avenue
New York, NY 10003
schober@newschool.edu

Abstract

An important question for some spatial language systems is how to select the frames of reference that underlie the production and interpretation of spatial expressions. The psychological experiments described here examine whether the usual classification scheme for reference frames should be modified to include a reference frame for the addressee of a spatial expression, or whether considering the addressee to be like another object in the environment is appropriate. The experiments compare how quickly people can describe the same location from the perspective of another person and from the perspective of an object (a chair). Results suggest that taking an object's perspective can be easier than taking another person's. This has implications for (and raises questions about) how systems should represent conversational partners and how often they should instantiate reference frames during processing of spatial expressions.

Introduction

Many spatial descriptions cannot be produced or interpreted correctly without being assigned a reference frame. For example, "on the left" might refer either to a speaker's left, her addressee's left, or the left of an object in the environment like a car or a chair. The exact spatial relation cannot be inferred without knowing which frame of reference is being used. This is well known to AI theorists and system designers, who regularly include the frame of reference as a necessary component for computations involving spatial language (see, e.g., Gapp 1995; Herskovits 1986; Lang, Carstensen & Simmons 1991, among many others).

But computational theories tend to focus on details of the production or interpretation of spatial terms once a reference frame has already been selected. Less attention has been paid to how frames of reference should be selected in the first place.

For many current applications, this is the appropriate strategy, because reference frame selection isn't a problem for them. For example, if the system and user are guaranteed to have the same vantage point on a scene, assuming a viewer-centered frame of reference is safe. The user is likely to assume that the system shares her frame of reference, and to produce and interpret spatial descriptions accordingly.

But if systems applications are to reflect the full repertoire of human spatial language use, reference frame selection will become a much more complicated issue. This will be especially true in any case where the originator of a spatial description (system or user) does not share the vantage point of the addressee (system or user). This could happen, for instance, in a multi-user spatial application where different communicating users have different points of view. Or it could happen when a mobile user communicates with an off-site system about a real-world scene (see, for example, Davis' 1989 Back Seat Driver system, in which the reference frame of the user, a car driver, must be recomputed as the driver's vantage point and distance from other objects change). Multiple coordinate systems will need to be represented in such systems, and there will be reference frame selection issues in both production and interpretation of spatial language.

For such applications it will be important to understand more about how people choose reference frames in conversations with other people, and what they deem their choices to be. This is not because systems must be modeled on human processing strategies; they needn't be, although sometimes this can be desirable. Rather, the usability of intelligent interfaces

depends critically on how well the interface designers have been able to estimate human cognitive and linguistic behaviors.

In earlier work on human conversations about locations (Schober 1993, 1995) I have proposed that one major factor in reference frame selection is *effort*. Different levels of effort are likely to be associated with taking different perspectives. I have also proposed that people have more options in choosing reference frames than theorists usually suppose.

Here I briefly review these proposals, and then describe some further studies (Schober & Bloom 1995) that test one aspect of them. Then I discuss unresolved issues raised by the results and potential computational implications.

Classification of reference frames

The standard way to classify reference frames is as *deictic, intrinsic,* and *extrinsic* (see, for example, Carlson-Radvansky & Irwin 1993; Herskovits 1986; Levelt 1989; Retz-Schmidt 1988, among many others). The same spatial description can express different locations depending on which of these frames it reflects. Deictic expressions reflect the speaker's or viewer's perspective; they have also been called egocentric and speaker-centered. Intrinsic expressions reflect the intrinsic tops, bottoms, lefts, rights, fronts and backs of objects in an environment that have these features (lamps have tops and bottoms, but no lefts or rights; chairs have all these features). Intrinsic expressions have also been called object-centered. Extrinsic expressions reflect unchanging environmental attributes like gravity and compass points. So "up" means "away from the pull of gravity" regardless of the frames of reference of people or objects in the environment. Extrinsic expressions have also been called absolute and environment-centered.

There are problems with this scheme. One problem has to do with cross-linguistic validity: As Levinson (1996) points out, some languages do not include all these options. Another problem is that some spatial descriptions can be produced and interpreted without reference to any of these coordinate systems. These include expressions like "between," "near," and "in the middle," which Levelt (1989) calls *local references without a coordinate system* and I have called *neutral* with respect to frames of reference (Schober 1995). Another problem is that some expressions are ambiguous with respect to coordinate systems. That is, when a speaker's left and right coincide with the intrinsic left and right of an object in a scene, an expression that uses "left" or "right" can't be clearly marked as indicating the speaker's or the object's frame of reference.

The problem I will be concerned with here has to do with the classification of *addressees*--that is, conversational partners that speakers are currently addressing. (Addressees contrast with other conversational participants who are not currently being addressed and with listeners like overhearers who are not part of the current conversation, see Schober & Clark 1989).

My simple observation is that in human spatial language use there is always a speaker (or writer) who originates the utterance. There is always an addressee or audience of some sort, whether it is a single person in the immediate environment who can provide feedback of understanding or a vaguer unspecified future audience. And there is always a reason for the spatial communication, like giving directions or localizing an object for some purpose. Spatial expressions do not arise in a vacuum, and a whole complex of contextual factors figures into their production and interpretation.

In the simplest two-person face-to-face communication, then, there are always *two* human coordinate systems involved from the start, no matter what frame of reference the speakers are using in their spatial descriptions. And these two coordinate systems can never coincide perfectly, although they can be close enough to be treated as identical for current purposes (see Herskovits 1986). For convenience I will call the first person the speaker and the second person the addressee. The standard scheme classifies the addressee's frame of reference as either deictic or intrinsic, and researchers have done both.

Those who classify the addressee's perspective as deictic (e.g., Herskovits 1986; Retz-Schmidt 1988) rely on the fact that speakers and addressees are both human observers of scenes. Those who classify the addressee's vantage point as intrinsic. classify *all* non-egocentric descriptions as intrinsic (e.g., Fillmore 1982; Levelt 1989), and this includes addressees. Addressees are just like other objects in a speaker's environment that have their own fronts, backs, lefts, rights, tops and bottoms.

In a sense, both characterizations are right, and in a sense they are both wrong. Addressees do share the critical property with other objects in the environment of having their own coordinate systems that speakers can refer to. And in taking an addressee's or an object's perspective, a speaker is being non-egocentric by taking the point of view of an entity other than herself.

But addressees are more like speakers than other objects in the environment. Speakers must coordinate frames of reference with their conversational partners, but they don't have to take the needs of inanimate objects into account. And spatial descriptions are part of purposeful communication with addressees, not objects.

Another way addressees are more like speakers is that there are some kinds of spatial descriptions unique to human coordinate systems: the kind that project onto the environment in the visual field. For example, a description like "on the left" can be uttered and interpreted with respect to the speaker's own intrinsic left, as when the target object whose location is being described is on the speaker's immediate left. Or it can be uttered and interpreted as a projection of the speaker's intrinsic left onto the environment, as when the target object is in the left half of the speaker's visual field. Speakers can utter these "projected" kinds of expressions from the addressee's point of view, describing what is on, for instance, the left half of the addressee's visual field. But these kinds of descriptions seem less appropriate for objects that don't have visual fields, like chairs. It seems odd to describe something in front and to the left of a chair as merely being on the chair's left.

My proposal (Schober 1993, 1995) is that addressee-centered descriptions reflect a full-fledged perspective in their own

right. This perspective shares features with both speaker- and object-centered perspectives, but it is qualitatively different. This proposal requires empirical validation. One important kind of empirical validation would be a demonstration that speakers treat addressees differently from other objects in the environment during the production and interpretation of spatial descriptions. In the studies described here (Schober & Bloom 1995), we use a standard cognitive psychological method to examine this in the case of production. If people take different amounts of time to produce object- and addressee-centered descriptions, this suggests that the mental processes involved in their production are different. It also lends plausibility to the proposal that addressee-centered descriptions should be classified differently than object-centered descriptions, and that addressee-centered descriptions reflect a frame of reference in their own right.

Effort and reference frame selection

Effort seems to be one of the factors involved in reference frame selection (Schober 1993, 1995). Different frames of reference require different amounts of effort to produce and comprehend. The evidence from earlier studies of speakers' spatial descriptions (e.g., Herrmann et al. 1987) shows that speakers find it easier to take their own perspectives than the perspectives of their conversational partners, as measured by latency to production of locative expressions. Carlson-Radvansky & Irwin (1994) have shown that verifying a locative expression like "above" is easier when it is true from multiple frames of reference, rather than only one.

But the role of effort in reference frame selection is complex. People don't produce only descriptions that are easiest to produce (that is, egocentric descriptions). And they don't automatically interpret the spatial expressions they hear as reflecting the frame of reference that would be easiest for them to understand (their own frame of reference). I propose that rather than minimizing their own effort, people make much more complex judgements. When they hear a spatial expression, they take into account the effort their conversational partner would have to make in producing that expression from that viewpoint. When they produce a spatial expression, they take into account not only how hard they themselves will find it to produce the expression, but also how hard their addressee will find the expression to interpret. That is, while speaker-centered utterances are easiest for speakers to produce, they are harder for addressees to understand than addressee-centered utterances.

In two laboratory studies (Schober, 1993, 1995) I have investigated the perspective choices conversational participants make as they describe spatial locations to each other. In the first, speakers described a simple display, where their descriptions could be either speaker-centered, addressee-centered, or ambiguous (in cases where the speaker's and addressee's viewpoints coincided, the expressions themselves could have reflected either perspective). In this study, speakers averaged far more addressee-centered than speaker-centered utterances. So on average speakers minimized effort for their partners by taking the addressee's perspective There was, however, substantial variability in the strategies pairs chose, and in some pairs both parties spoke completely egocentrically. (In my laboratory we are currently following up on this by investigating how individual differences in perspective-taking strategies relate to underlying mental rotation abilities).

In the second study, speakers describing a far more complex display avoided speaker-centered utterances almost entirely. For this display, descriptions could either be speaker-centered, addressee-centered, ambiguous, object-centered, or neutral with respect to reference frames. Again, speakers on average preferred to take their partner's perspective over their own. But even more popular were neutral descriptions, and these neutral descriptions increased in frequency over the course of the conversations. Neutral descriptions don't require either party to shift perspectives away from their own, because they don't reflect a coordinate system at all. So in a sense, what speakers in this study were doing was minimizing effort not only for their partners but, to the extent they could, for themselves. Their reference frame selection minimized effort for both parties.

In this second study, object-centered perspectives were avoided almost completely. But because the objects in the display didn't have very compelling intrinsic parts, we shouldn't make too much of this avoidance. Miller & Johnson-Laird (1976) claim that taking an object's perspective is less difficult than taking another person's. Thus object-centered descriptions should be preferred, because they requires an intermediate level of effort for both parties. That is, taking the object's perspective is harder for the speaker than speaking egocentrically, but easier than taking the addressee's perspective. Understanding a description from the object's perspective is harder for the addressee than understanding an addressee-centered description, but easier than understanding a speaker-centered perspective.

In the studies I describe here, we examine Miller & Johnson-Laird's claim using a display that contains an object with compelling intrinsic parts. If Miller & Johnson-Laird are right, then describing a location from the object's perspective should be harder than speaking egocentrically, but easier than taking the addressee's perspective. This is by no means necessarily true. It could be just as difficult to take the perspective of an object as to take the perspective of the addressee, since both require shifting points of view away from one's own. Or it could be easier to take the addressee's point of view than an object's, because spatial expressions are designed for particular addressees and not objects.

The studies

In both studies, we adapted Herrmann et al.'s (1987) method, in which speakers describe locations on simple computer graphics displays from their own or from other perspectives, and latencies to production are measured. Speakers viewed color displays of 3-D rooms on a CRT screen, and we measured latencies to production of locative expressions as they described the location of a ball relative to a person (who represents the addressee) or a chair (an object that has a front, back, left and right). Speakers viewed exactly the same scenes on different occasions, but they were required to take different perspectives on the different viewings. Thus we could collect

within-subjects comparisons of latencies to production.

Experiment 1

In the first study, 19 participants viewed a set of color displays of a 3-D room (see Figure 1 for sample displays). Each display contained a chair, a human figure, and a ball (the *target object*). For each display, participants were asked to describe the location of the target object using only the words *front, back, left,* or *right.*

The participants did this three times for the entire set of displays: once from their own perspective, once from the chair's perspective, and once from the person's perspective. The ordering of displays was randomized and the ordering of perspectives counterbalanced.

On each of the 32 displays, the positions and orientations of the chair, person and ball varied as follows: The **chair**, always in the center of the room, faced either to the viewer's left or right (90° or 270°), or toward the viewer (180°). The **person**, always against a wall, faced into the room either nearly at the same vantage point as the viewer (0°), on the left or right wall (90° or 270°), or on the far wall (180°). The **ball** was placed so that the same spatial description (*left, right, front,* or *back*) could not be correctly used to describe its location from more than one perspective. For example, if the ball was to the viewer's left, it would not simultaneously be to the left of either the chair or the human figure in the display. This was (1) so we could be sure that participants were using the requested perspectives, and (2) so we could be sure that only one frame of reference was being activated at a time. Spatial descriptions may be easier to verify when they are true from more than one perspective (Carlson-Radvansky & Irwin, 1994, 1995), and they may also be easier to produce.

For each trial, participants would press a key and an image of the room, chair, and person would appear. After pressing a key again, there was an audible beep, and the ball appeared somewhere in the room. The participant would then describe the ball's location using the requested perspective.

Reaction time results showed that speakers reliably found it easiest to take their own point of view. This replicates Herrmann et al.'s (1987) results. Speakers found it harder to take the chair's perspective, and hardest to take the other person's perspective. People found it easier to take the addressee's and object's perspective at 90° and 270° offsets than at the 180° offset.

People were very slow to take the perspective of the addressee when it shared their vantage point (speakers looked over the shoulder of the computer-graphic person in the display). We don't believe this is a substantive finding, since our speakers seemed to find these displays confusing. As one would expect, when participants took the addressee's perspective, the chair's offset had no effect. When they took the chair's perspective, the addressee's offset had no effect.

Under certain circumstances, then, the processes involved in taking the perspective of another person seem to differ from those involved in taking the perspective of some other object in the environment. This provides initial support for my proposal that addressee-centered descriptions reflect a frame of reference in their own right. It also neatly supports Miller & Johnson-Laird's (1976) claim that object-centered descriptions are of intermediate difficulty for speakers--harder than speaking egocentrically, but easier than taking the addressee's perspective.

But there were potential concerns with Experiment 1. First, the addressee was always against the wall and the chair was always in the center of the room. Since chair orientation and addressee position varied from trial to trial, the addressee's position changed significantly and quickly, while the chair's did not. Speakers may have found it harder to take the addressee's perspective in Experiment 1 merely because of this.

Second, for the particular stimuli used in the study, the task of taking the addressee's perspective had a somewhat different quality than the task of taking the chair's perspective. When the ball was on the chair's left, it was on the chair's immediate left, while when the ball was to the left of the human figure, it was in the left half of the human figure's field of vision, but also out in front. This could be why the speakers In Experiment 1 were slower to take the addressee's perspective.

Third, we weren't convinced that speakers in Experiment 1 really treated the human figure as an addressee. Perhaps they merely conceived of it as a more complicated object, like a statue of a person. If this was the case, then our results wouldn't show that addressee-centered descriptions reflect a distinct frame of reference. They would merely show that our speakers treated our CRT representation of the addressee as a more complicated object than the chair.

Experiment 2

To deal with these concerns, we ran Experiment 2. 16 participants followed exactly the same procedure as for Experiment 1, but with certain differences. Each of the 336 displays contained either the chair or the person, but not both. Both the chair and the person appeared in the center of the room and against the wall equally often, so that we could be find out whether our results from Experiment 1 merely reflected the fact that the chair and addressee had appeared in different positions in the room. Also to address this concern, on 12 of the 14 blocks of trials, the chair or the person appeared in the same predictable spot from display to display. On 2 of the 14 trials, they moved around as in Experiment 1.

To ensure that speakers really treated the representation of the addressee as an addressee, we told participants that another student who would see the displays from the human figure's perspective would be listening to their descriptions. That student would then be required to locate the ball on the basis of those descriptions.

We used only displays of situations where *left* or *right* would be appropriate, and we used only situations where the target object was on the chair's or person's immediate left or right. Thus we could ensure a fair comparison between addressee- and object-centered descriptions, since no descriptions would

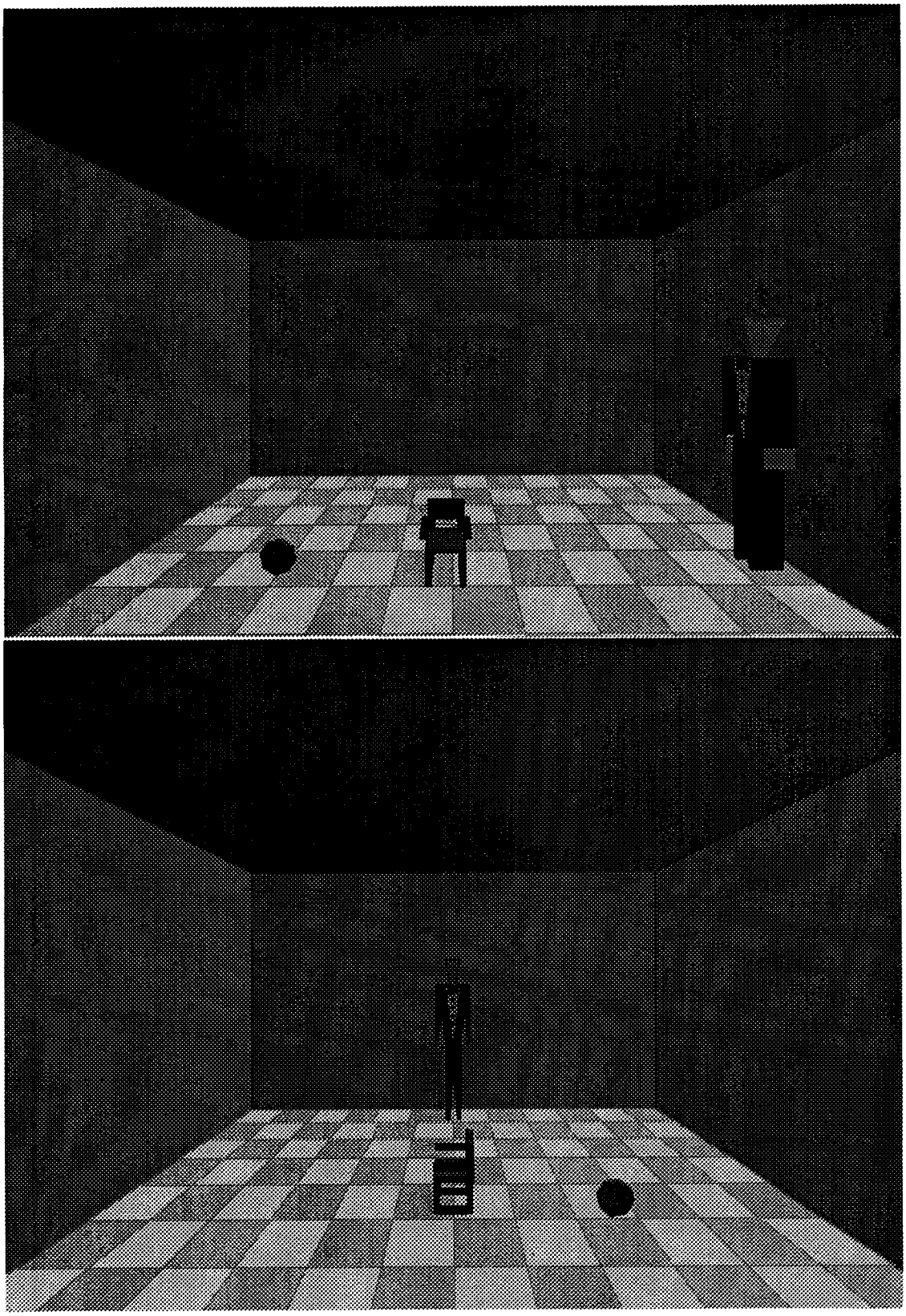

Figure 1: Sample displays, Experiment 1

be the sort "projected" into the environment. We decided it was reasonable to omit the *front* and *back* situations from Experiment 1 to avoid presenting the participants with an unreasonable number of stimuli. This was justified because in Experiment 1 participants had exactly the same pattern of results for producing *left* and *right* as for *front* and *back*. We also omitted 0-degree offset displays, because Experiment 1 participants had found these confusing. (See Figure 2 for sample displays).

The results replicated those of Experiment 1. Once again, addressee-centered descriptions took reliably longer than object-centered descriptions. This shows that the findings in Experiment 1 didn't result only from the problems with its stimuli.

People found it much easier to take either perspective when the addressee or object remained in a predictable spot. The difference between taking the addressee's and object's perspectives was marginally greater when the addressee or object moved from trial to trial. The differences were less extreme when the addressee or object remained in a predictable spot. But addressee-centered descriptions were nonetheless more difficult, whether the reference object (the chair or the person) moved from trial to trial or not.

Together, these experiments show that taking the spatial perspectives of another person is different than taking the perspective of a non-human object (a chair). The results support Miller & Johnson-Laird's (1976) proposal that taking an object's perspective is of intermediate difficulty; more difficult than speaking egocentrically, and less difficult than taking a conversational partner's perspective. The results also argue against classifying addressee-centered descriptions as either deictic or intrinsic. They seem to have special properties all their own.

Computational Implications

We begin to see that spatial perspective-taking in conversation is a complicated affair. The details of human perspective-taking behavior have implications for models of spatial language, and here are some of them. Just how serious these implications are depends, of course, on the application.

1. If the addressee-centered frame of reference really is a reference frame in its own right, then systems where addressees (system or user) have different vantage points than speakers (user or system) will have to include an addressee-centered reference frame in its repertoire. Users speaking to systems should be understood whether they speak egocentrically or take an addressee-centered perspective. Systems involving multiple users with different vantage points will have to represent the points of view of each user and interpret utterances in light of those reference frames.

2. If an effort-based account of perspective choice is right, the data described here suggest that taking an object's perspective can be easier than taking another person's perspective. These data do not show that people prefer taking an object-centered perspective to an addressee-centered one, since in our experiments the participants were required to speak from the chair's or the person's perspective repeatedly for a block of displays. But they do suggest that in situations where speakers choose to (or must) minimize effort, object-centered descriptions will be more difficult for them to produce than egocentric ones, but less difficult than addressee-centered ones. This may ultimately affect how they choose reference frames.

3. Exactly how are conversational partners mentally represented? Are they like other objects but more complex, or are they special? The results here present very preliminary evidence that addressees are mentally represented in a special way. This is consistent with Clark & Marshall's (1981) proposal that each particular addressee is distinctly represented, and that information about particular addressees is active and salient during language processing with those addressees. But if this is so, it raises a puzzle. If addressees are really specially available during utterance production, then why should speakers take *longer* to produce addressee-centered descriptions than object-centered descriptions? It seems reasonable that there should be extra processes involved in addressee-centered description, and these should take extra time (Herrmann 1989). But shouldn't the addressee be available during object-centered descriptions as well, since the utterances are presumably designed for the addressee? And so shouldn't object-centered descriptions take even longer than addressee-centered descriptions, because both the object's perspective *and* the addressee's perspective must become active?

Our data don't provide an answer to these questions. One possibility is that in taking an object's perspective the addressee's point of view doesn't have to be considered, and thus it doesn't have to be fully activated.

4. If effort is a significant component of reference frame selection, then an important issue will be how to represent it. Wherever in a system reference frame selection happens, there will need to be a component that weights particular reference frames for how difficult they are to produce and interpret in the current situation. This weighting will have to take into account contextual factors like particular arrangements of objects in scenes and particular displacements of speaker and addressee vantage points. A further complication to this picture is that different people may have different weightings for effort. A study I am currently engaged in analyzing is suggesting that people's mental rotation abilities affect their strategies in selecting perspectives, in combination with their assessments of their conversational partners' abilities. If this is the case, then user models may need to represent a component that weights reference frame likelihood for particular individuals.

5. One provocative piece of data here is that the difference between taking the addressee's and object's perspectives was less extreme when the addressee or object remained in a predictable spot. The appropriateness of generalizing this result beyond the studies described here depends on whether spatial descriptions in the situations to be modeled are more like the experimental condition in which the reference object moved from trial to trial or the stationary case. Both situations can happen in the real world, although our experimental situation was artificial in the regularity and predictability with

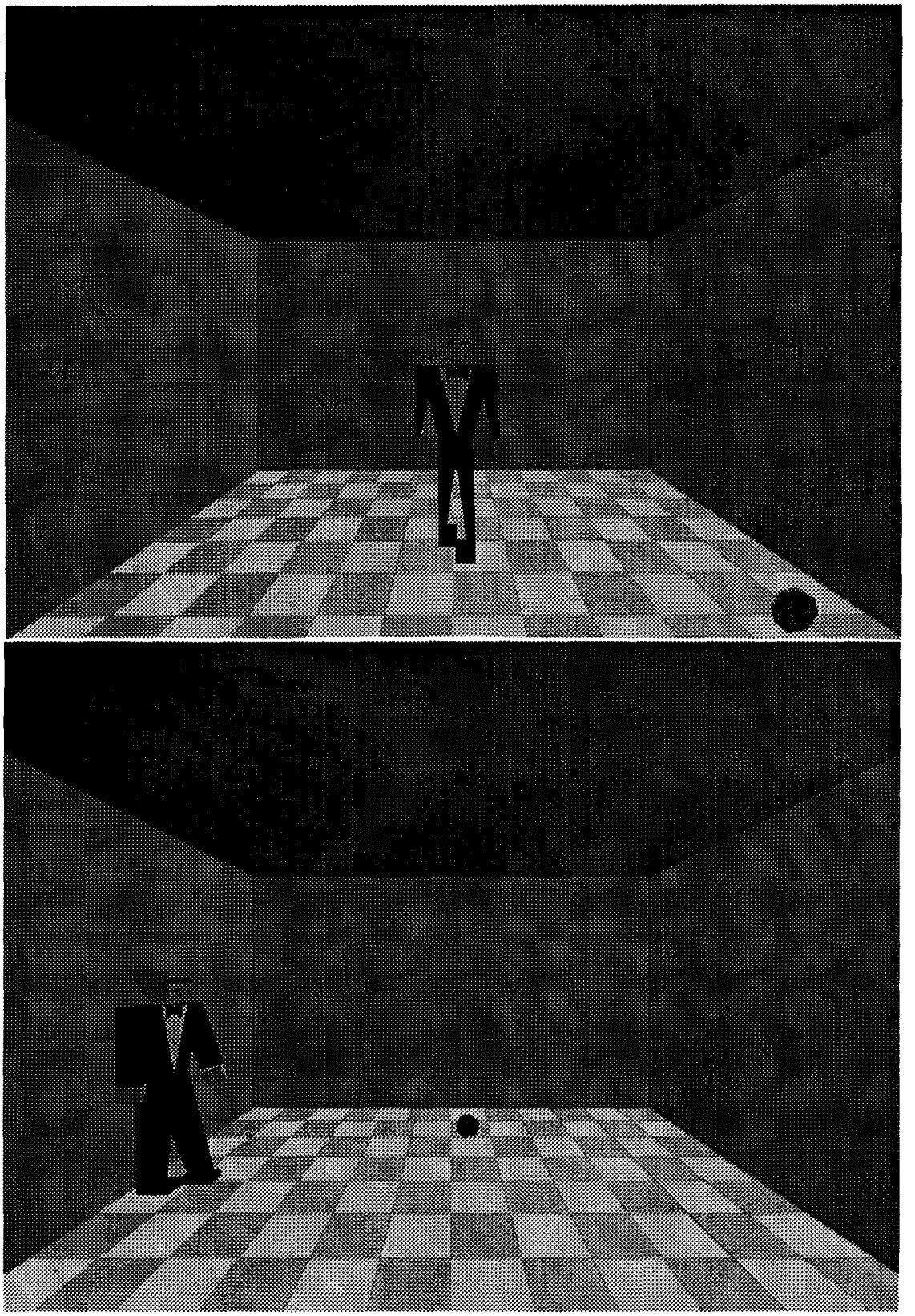

Figure 2: Sample displays, Experiment 2

which experimental participants had to produce their descriptions.

These studies also raise questions for the future that have significant implications for systems that try to model changing environments. When do speakers instantiate a reference frame or coordinate system? Certainly they must do it before they produce their spatial expressions (Herskovits, 1986). But how often do they update them? Once a reference frame is instantiated is it kept until speakers get evidence that circumstances have changed? Or do speakers reconfirm reference frames regularly, especially if intervening spoken material hasn't kept a reference frame active? One speculation consistent with this result is that speakers instantiate a frame of reference for a first spatial description, and it can be more difficult for them to do this for an addressee than for an object (at least the one we used). But once the frame of reference has been instantiated, it needn't be reinstantiated for subsequent descriptions, as long as the reference object or addressee doesn't change vantage points.

Finally, I should note that effort is not the only factor involved in reference frame selection. Speakers are affected by the reference frames they have already chosen and the reference frames their conversational partners have chosen (see Garrod & Anderson 1987). Speakers are also affected by what reference frames their conversational partners give them evidence of having understood, and they will reframe their descriptions from a different perspective when they are misunderstood (Schober 1993). This should remind us that in human conversation repair systems act as a safety net for misunderstanding, and it is a reasonable strategy in modeling spatial language to invest significant computational resources in the diagnosis and repair of misunderstanding (Brennan & Hulteen 1995; Schober 1996).

References

Brennan, S.E., and Hulteen, E.A. 1995. Interaction and Feedback in a Spoken Language System: A Theoretical Framework. *Knowledge-Based Systems* 8:143-151.

Carlson-Radvansky, L.A., and Irwin, D.E. 1993. Frames of Reference in Vision and Language: Where is Above? *Cognition* 46:223-244.

Carlson-Radvansky, L.A., and Irwin, D.E. 1994. Reference Frame Activation During Spatial Term Assignment. *Journal of Memory and Language*, 33, 646-671.

Carlson-Radvansky, L.A. 1995. Cooperation Among Spatial Reference Frames. Poster session presented at the 36th Annual Meeting of the Psychonomic Society, Los Angeles, CA.

Clark, H.H., and Marshall, C.R. 1981. Definite Reference and Mutual Knowledge. In A.H. Joshi, B. Webber, and I.A. Sag, eds., *Elements of Discourse Understanding* (pp. 10-63). Cambridge, UK: Cambridge University Press.

Davis, J.R. 1989. Back Seat Driver: Voice Assisted Automobile Navigation. M.Sc. Thesis, Dept. Of Architecture, MIT.

Fillmore, C.J. 1982. Towards a Descriptive Framework for Spatial Deixis. In R.J. Jarvella & W. Klein, eds., *Speech, Place, and Action* (pp. 31-59). Chichester: Wiley.

Gapp, K.-P. 1995. Object Localization: Selection of Optimal Reference Objects. In Proceedings of the 2nd International Conference on Spatial Information Theory, COSIT 95.

Garrod, S., and Anderson, A. 1987. Saying What You Mean in Dialogue: A Study in Conceptual and Semantic Co-ordination. *Cognition* 27:181-218.

Herrmann, T., Bürkle, B., and Nirmaier, H. 1987. Zur hörerbezogenen Raumreferenz: Hörerposition und Lokalisationsaufwand [On Listener-Oriented Spatial Reference: Listener Position and Localization Effort]. *Sprache & Kognition* 3:126-137.

Herrmann, T. 1989. Partnerbezogenen Objektlocalisation--ein Neues Sprachpsychologisches Forschungsthema [Partner-oriented Localization of objects--a New Psycholinguistic Research Topic], Bericht Nr. 25, Forschergruppe "Sprechen und Sprachverstehen im sozialen Kontext," University of Mannheim.

Herskovits, A. 1986. *Language and Spatial Cognition: An Interdisciplinary Study of the Prepositions in English.* Cambridge, UK: Cambridge University Press.

Lang, E., Carstensen, K.-U., and Simmons, G. 1991. *Modeling Spatial Knowledge on a Linguistic Basis* (Lecture Notes in Computer Science 481). Berlin: Springer-Verlag.

Levelt, W.J.M. 1989. *Speaking: From Intention to Articulation.* Cambridge, MA: MIT Press.

Levinson, S. 1996. Frames of Reference and Molyneux's Question: Cross-Linguistic Evidence. In P. Bloom, M.A. Peterson, L. Nadel, and M. Garrett, eds., *Space and Language*. Cambridge: MIT Press. Forthcoming.

Miller, G.A., and Johnson-Laird, P.N. 1976. *Language and Perception.* Cambridge, MA: Harvard University Press.

Retz-Schmidt, G. 1988. Various Views on Spatial Prepositions. *AI Magazine* 9(2):95-105.

Schober, M.F. 1993. Spatial Perspective-Taking in Conversation. *Cognition* 47:1-24.

Schober, M.F. 1995. Speakers, Addressees, and Frames of Reference: Whose Effort is Minimized in Conversations About Locations? *Discourse Processes* 20:219-247.

Schober, M.F. 1996. How Addressees Affect Spatial Perspective Choice in Dialogue. In P.L. Olivier and K.-P. Gap, eds., *Representation and Processing of Spatial*

Expressions. Hillsdale, NJ: Lawrence Erlbaum. Forthcoming.

Schober, M.F., and Bloom, J.E. 1995. The Relative Ease of Producing Egocentric, Addressee-Centered, and Object-Centered Spatial Descriptions. Poster session presented at the 36th Annual Meeting of the Psychonomic Society, Los Angeles, CA.

Schober, M.F., and Clark, H.H. 1989. Understanding by Addressees and Overhearers. *Cognitive Psychology* 21:211-232.