Running head: DESCRIPTIONS OF SIMPLE SPATIAL SCENES

Descriptions of Simple Spatial Scenes in English and Japanese

Scott D. Mainwaring | Barbara Tversky | Motoko Ohgishi

Interval Research Corp. | Stanford University | Seibo Jogakuin Junior College

Palo Alto, California | | Kyoto, Japan

Diane J. Schiano

Interval Research Corp.

Palo Alto, California

Correspondence concerning this article should be addressed to Barbara Tversky, Psychology Department, Jordan Hall (Building 01-420), Stanford, CA 94305-2130; telephone +1 650 725-2440; fax +1 650 725-5699; email bt@psych.stanford.edu

Keywords: spatial cognition; locative expressions; perspective taking; reference frames; social effects; cultural differences

Abstract

How do people describe the location of a target object to another? This task requires a reference object or frame and terms of reference. Traditional linguistic analyses have loosely organized perspectives around people, objects, or environments as reference objects, using reference terms based on a viewpoint or the intrinsic sides of an object, such as *left*, *right*, *front*, and *back* or based on the environment, such as *north*, *south*, *east*, and *west*. In actual communication, social, spatial, and cognitive factors may also affect perspective choice. We examine those factors by varying the spatial information (landmarks and cardinal directions), the communication task (relative cognitive burden to speakers and addressees), and the culture of participants (American and Japanese). Speakers used addressees' perspectives more when addressees had the greater cognitive burden. They also used landmarks and cardinal directions when they were available, especially to avoid difficult discriminations like left/right. Some cases appearing to be perspective taking can be interpreted as using a person as a landmark. Finally, terms like *near* indicating close proximity were preferred to *far* and to terms requiring projection of directions. Globally, perspective choices of American and Japanese samples were strikingly similar; that is, Japanese did not select addressees' perspectives more than Americans. The traditional linguistic analyses need to be enhanced to account for effects of cognitive, situational, and social factors.

Descriptions of Simple Spatial Scenes

One situation we cannot escape is the spatial environment in which we conduct the business and pleasure of life. Often, we are called on to describe that environment, for example, to inform others where something or some place is located. When we observe objects in the world, we are necessarily bound by our own perspective, by the point in space of our sense organs in our bodies. Language and thought free us to take other perspectives. In fact, conceiving of perspectives other than our own fundamental for such different and essential cognitive skills as recognizing objects (e.g., Ullman, 1996), wayfinding (e.g., Tversky, 1996), and communicating with others (e.g., Clark, 1996). Spatial language is useful not just for conveying spatial information but also for conveying abstract concepts (e.g., through spatial metaphor, see Lakoff & Johnson, 1980). Languages typically provide many options for describing space (see, for example, papers collected by Bloom, Peterson, Nadel, & Garrett, 1996). Here, we explore some of the spatial, cognitive, and social factors determining selection of spatial descriptions.

The components of a spatial description include a target object (*figure* in Talmy's (1983) terms), a reference object or reference frame (*ground* in Talmy's terms), and terms of reference. These components are not necessarily independent. Often they are correlated and organized into what have been called perspectives. Reference objects or frames include a person in the scene, another object or landmark in the scene, or an environmental framework, such as the surrounding building or the cardinal directions. Terms of reference vary across languages. In many languages, including English and Japanese, there are at least three common sets of reference terms, those depending on a viewpoint and those depending on the intrinsic sides of an object, both of which use *left*, *right*, *front*, and *back*; and those depending on an environment, such as *ceiling*, *floor*, *entrance*, and *exit*, or *north*, *south*, *east*, and *west*.

A number of linguists and psychologists have analyzed the types of reference objects and reference terms used to describe locations in space in order to develop a systematic set of reference frames or perspectives (for example, Buhler, 1934/1982; Clark, 1973; Fillmore, 1975, 1982; Levelt, 1984, 1989, 1996; Miller & Johnson-Laird, 1976; Levinson, 1996; Talmy, 1983; Taylor & Tversky, 1996). One might expect that these analyses would find that different types of frames of reference correspond in a straightforward way with different types of reference origins/reference objects; i.e., that the so-called deictic (viewpoint dependent) frame of reference would be centered on a person, the intrinsic on an object, and the extrinsic on the environment.

However, as Levinson (1996) and others have observed, the alignment of origins, reference objects, and terms of reference leads to inconsistencies and confusions. In particular, intrinsic reference frames based on objects are formally the same as intrinsic reference frames based on people. This led Levinson to distinguish three types of reference frames, relative, intrinsic, and absolute, varying on origin, number, and type of terms of relation, and several other features (Levinson, 1996). Let us characterize the three systems using the display in Figure 1, consisting of an observer at the threshold of a zoo, facing north, in sight of an elephant and a tree. The location of the tree would be conveyed differently by each of the reference frames. The relative system has the ego (of the observer) as an origin and takes a three-termed relation. Using this relative system, the tree’s location would be described as "the tree is to the left of the elephant." The intrinsic system has its' origin in an object that has conventional sides and takes a two-termed relation. To locate the tree intrinsically in the scene, one would say, "the tree is in front of the elephant." Note that the location of the elephant relative to the tree could not be described intrinsically as the tree does not have intrinsic front, back, left, and right. Finally, the absolute system is environment-centered and uses a two-term relation, as in "the tree is west of the elephant."

Levinson's analysis not only clarifies the issues raised by previous analyses, but also is consistent with some spontaneous descriptions of space (Taylor & Tversky, 1996). People asked to describe environments they know well, such as a dormitory room, a convention center, or a small town, adopt one of three perspectives: a gaze tour, corresponding to standing at an entrance and scanning the environment (Ehrich & Koster, 1983); a route tour, corresponding to moving through the environment; or a survey, corresponding to scanning the environment from above.

For the most part, gaze descriptions use a relative frame of reference, route descriptions use an intrinsic frame of reference, and survey descriptions use an absolute frame of reference. Levinson's tripartite distinction was not meant to exhaust the possibilities for spatial descriptions; it was meant to organize the three common reference systems. In fact, actual descriptions contain more varieties of spatial descriptions. For one thing, in usage, origin objects and reference terms don't always align. In Taylor & Tversky’s (1996) corpus, route descriptions, based on an intrinsic frame of reference, were typically centered on a person ("you"), rather than an object. For another thing, many of the descriptions, nearly the majority, were mixed, that is, they switched perspectives, sometimes mid-sentence or mid-phrase. For example, intrinsic descriptions often used terms like north and south rather than terms like right and left. Many other kinds of reference terms also appeared in abundance in the spontaneous descriptions, terms like *near* and *besides*. Language in practice is not as neatly organized as language in theory.

There is yet another central aspect of language use that the various schemes for reference frames do not reflect. Language is typically used in a social context, one person speaking or writing to another. In describing space in a social situation, whose perspective is adopted, that of the speaker or the addressee or neither, is also significant (Schober, 1993; Tversky, 1996). None of the schemes developed so far have whose perspective as a central factor.

The importance of whose perspective is illustrated in the two-object task developed by Schober (1993) and adapted for the present research. In the two-object task, a speaker must distinguish to an addressee, though verbal means only, the location of one of two identical objects. For example, consider the task scenario from one of our experiments depicted in Figure 2. The task requests a minimalist spatial description, in the context of a cover story about a pair of cooperating spies and a means of covert communication. The scene includes two identical objects and two viewpoints, that of the participant or speaker, Agent U, and that of the addressee, Agent Z, the participant's partner in secret missions. The participant knows which of the objects is the target (the one containing a bomb, secret information, stolen jewelry, etc). The participant’s task is to tell Z the location of the target by writing a brief description into an

"Encoder Pad." In the minimal baseline scene depicted in Figure 2, there are limited ways to describe the location of the target, indicated by an arrow. The participant can adopt the participant's own perspective, saying "it's the one directly in front of me," or can adopt Z's perspective, saying "it's the one on your left."

Schober's (1993) task was designed to allow only two possibilities, using the speaker's perspective or the addressee's perspective. This important communication choice had not been discussed in the linguistic analysis of perspective. Schober found that most of the time, speakers took addressees' perspectives. Interestingly, the tendency to take the perspective of the addressee was stronger when participants recorded a message for a hypothetical addressee (who thus could not request clarification) than when they spoke interactively with an actual addressee. In a similar task, Herrmann (1989; cited in Schober, 1993) also found that speakers tended to take the perspective of their addressees. As Schober explains, there are several intertwined reasons why speakers may take their addressees' perspectives. For one thing, it seems more polite to recognize the addressee's perspective in directing a message to an addressee (Brown and Levinson, 1987). For another, it seems that constructing a message from the addressee's viewpoint would increase the likelihood that the message is understood. This is especially true when the speaker has the critical information so the cognitive burden of the addressee is greater. The speaker perceives the scene directly and knows which of the two objects is the critical one. In contrast, the addressee needs to understand the language of the speaker's statement and use that language to determine which of the two identical objects is critical.

In this project, we have enriched the two-object task in three ways, situationally, socially, and cognitively, to enlarge the exploration of choice of spatial description. Our first addition was to enhance the environments by providing landmarks or cardinal directions or both for some scenarios. In typical environments, there are more options than just speaker and addressee perspective, in particular, landmarks and extrinsic relations. In fact, a number of language communities do not use relative or intrinsic reference frames based on left and right; instead, they use extrinsic reference frames based on the cardinal directions (Levinson, 1996). Research on "cognitive maps" has shown that spatial knowledge is organized in large part around landmarks and environmental reference frames (e.g., Tversky, 1992), so there is good reason to believe that speakers would use them in describing spatial locations. Landmarks and cardinal directions differ in their salience in environments. Landmarks are typically visible, but cardinal directions often must be inferred from correlated features in the environment. Cardinal directions and landmarks also differ in the kind of spatial description they entail. Cardinal directions invoke an orthogonal reference frame, north-south east-west. To use them requires projecting directions onto space. Landmarks can be used with an orthogonal reference frame (intrinsic or extrinsic); however, they can also be used with simpler reference terms, such as *next to*, *near*, or *besides* that do not require projecting directions into space. What's more, there seem to be some reliable individual differences around choice of reference frame. In particular, it is said that men are more likely to use cardinal directions and women to prefer to organize spatial information around landmarks (e.g., Bever, 1992; Galea & Kimura, 1992). Finally, the presence of landmarks and an extrinsic reference frame allow speakers to take perspectives other than those of speaker and addressee. To investigate choice of personal perspectives, we observe spatial descriptions in the simple two-object task under different conditions described below. To investigate choices among personal perspectives, landmarks, and cardinal directions, we include scenarios with those as well as the two objects. Finally, to investigate relative preferences for landmarks and cardinal directions, we include some scenarios with only those as possibilities (remote scenarios).

Our second addition to the two-object task was to elicit utterances from both American and Japanese students. This cross-cultural addition is interesting because it provides another test of politeness and a test for cultural robustness at the same time. Japanese are regarded as an especially polite and deferential society, so to the extent that politeness is an issue in perspective choice, Japanese participants should be more inclined toward the perspective of others. However, to the extent that the patterns of spatial description are similar in American and Japanese participants, the general cross-cultural cognitive and social determinants of spatial description choice will be supported.

Our third addition to the two-object task was to vary the cognitive load of speakers and addressees by varying the task to be performed. If politeness is the major determinant of perspective, variations in cognitive load should have no effect on choice of spatial description. If cognitive factors are also operative, then choices should vary with relative cognitive burden. In Schober's task, only speakers knew which of the two objects was the critical one, and relayed that information to their addressees. Speakers got to choose the spatial description. Of course, we can assume that speakers want to be understood, so within their own limits and abilities, they will construct a spatial description with their addressees in mind (cf. Clark & Wilkes-Gibbs, 1986). In this case, addressees had the more burdensome cognitive task. If speakers realize this, they may ease the cognitive burden of their addressees by taking the addressees' perspective on the scene rather than their own, for example, saying "the one on your right" instead of "the one on my left." An utterance with the addressee's perspective should be easier for the addressee to understand than an utterance using the speaker's perspective. Easing the cognitive burden of their addressees seems to be one reason why speakers generally took their addressees' perspectives in Schober's task (Schober, 1995).

Relative cognitive burden of speakers and addresses has many incommensurate and interwoven determinants. One is the information per se, who has it, the speaker or the addressee? Knowing the requisite information, especially in direct perception, is less burdensome than having to obtain it, especially through language rather than direct perception. A second factor is the spatial situation, the array of objects and the locations of speaker and addressee with respect to each other and to the objects. The easiest case, hence not considered here, is when speaker and addressee share the same point of view. When speaker and addressee do not share the same point of view -- when they are facing each other or at right angles, for example -- communicating the target object is more difficult. The difficulty of communicating the target object can be complicated by how the array is aligned, with the speaker or the addressee or both. For example, some alignments require the use of the terms, *left* or *right*, terms that are more difficult to produce and comprehend than terms like *in front* (e.g., Bryant & Tversky, 1992). Communication difficulty is also affected by the presence of landmarks and/or cardinal directions. In the spatial arrays used here, landmarks do not require computing projecting directions in space, as personal perspectives and cardinal directions do. Computing and projecting directions in space requires cognitive effort. A third factor affecting relative cognitive burden of speakers and addressees is production and comprehension of utterances. As noted, the spatial array and relative viewpoints of speaker and addressee affect difficulty of producing and comprehending spatial descriptions. Also as noted, it is easier to translate a known spatial situation into language than to translate an utterance into an imagined spatial situation. Clearly, all the contributions to cognitive burden, critical information, spatial array, and language are interdependent, far from independent. For example, speakers construct utterances with the addressee's spatial and informational situation in mind.

The communicative situation introduced by Schober and adapted here is more restricted than the usual communication situation. The communications are one-way and one-time; there is no opportunity to create common ground through extended interaction, so no conclusions can be drawn about that. The situation is further altered by writing rather than speaking and by an imagined rather than actual addressee, so the terms *speaker* and *addressee* are extended from their literal use. These restrictions nevertheless allow a rich variety of spatial descriptions that can be studied as a function of variations in spatial situation and communicative context, our immediate research goals.

In the present project, we varied informational contributions to relative cognitive burden by varying the task of the speaker, the participant, across experiments. We have termed Schober's task the "tell other" task. In it, speakers have access to the information and need to construct an utterance for their addressees, who, in turn, must comprehend the utterance and use it to decide which of two identical objects is being referred to. Thus, addressees have a greater cognitive burden than speakers (Schober, 1992). We added two other tasks that vary the relative cognitive load between speaker and addressee. The first of these is the "ask other" situation.

Here, the addressee, rather than the speaker, has the relevant information, a relative advantage to the addressee. The speaker needs to find the information, but cannot ask directly. Instead, the speaker has to formulate a yes/no question that the addressee can answer directly from given perceptual information. This task increases the burden on the speaker and decreases it on the addressee, so that their cognitive burdens are more balanced than in the "tell other" case. If speakers take addressees' perspectives primarily to be polite, then participants should take addressees' perspectives equally often in the "tell other" and "ask other" situations. However, if speakers take addresses' perspectives because of considerations of relative cognitive burden of the communication partners, then speakers should take the perspectives of their addresses less often in the "ask other" than in the "tell other" situations. The final task we investigate is the "tell self" condition. In this case, speakers make a note for themselves to identify the critical object at a later time. Although the speakers' partner is in the scene, the speaker is communicating to him- or herself, not to the partner. In this case, speaker and addressee are the same, so if perspective choice depends solely on communication partners, the speaker's own perspective should always be selected.

To sum, if speakers are taking into account relative cognitive burden in choosing a perspective, they should choose their own perspective least in the "tell other" task, more in the "ask other" task, and even more in the "tell self" task (all other factors held constant). Politeness alone would dictate that speakers take their addressees' perspectives equally in the "tell other" and "ask other" conditions, and never in the "tell self" condition. Taking into account relative cognitive burden, like politeness, is a social factor, but it is deeply intertwined with a cognitive factor, and thus has different implications.

We varied the difficulty of the spatial array within each of the three information tasks in similar fashions. In a more complex task, Schober (1995) found that speakers were more likely to use neutral or speaker perspectives when speakers and addresses had different perspectives on the scene than when they shared the same perspective. In our task, speakers and addressees always had different perspectives on the scene: they either faced each other or were offset by 90 degrees. Difficulty of constructing a description was varied in another way. Left/right discriminations are more difficult than discriminating other spatial dimensions such as front/back (Franklin & Tversky, 1990). The expectation within task is the same as that between tasks. Speakers will attempt to ease the cognitive burden of their addressees when their addressees have greater relative cognitive burden as long as speakers are able to adopt their addressees' perspectives with relative ease. Thus, speakers should avoid using *left* and *right*, preferring, for example, landmarks or cardinal directions if possible. When using *left* or *right* is unavoidable, speakers should prefer addressees' left and right to the extent that the cognitive burden is greater for addressees, as long as speakers themselves are able to make that distinction.

Experiment 1: Tell Other

In Experiment 1, each scenario required the participant to produce for Agent Z a locative expression which would distinguish which of two identical objects concealed a target item. The scenarios also depicted the locations of participants and Agent Z, and in some cases, the location of a landmark or an indication of the direction of north.

Method

Participants. 71 undergraduates (37 male, 34 female) in introductory psychology courses at Stanford University and 70 undergraduates (35 male, 35 female) at Kanazawa University participated for credit towards fulfillment of course requirements. Data from one Stanford participant was excluded from analysis because she did not answer many (the last 13) of the 28 scenario questions.

Materials. 30-page questionnaire packets were assembled, consisting of a cover page, 28 scenario pages, and finally a page of background questions.

The cover page was in the form of a memo from the “Secret Operations Agency” to “Agent U” (the participant) explaining “your mission... is to help Agent Z accomplish various dangerous but essential tasks in a variety of environments.” Participants were told that for security reasons communication would be by “Encoder Pads” which would allow Agent Z “to send you a simple question and you to answer with a single short message.” Agent Z would not be able to respond to ask for clarification so they should "make each of your messages clear and concise." Participants were informed that the information available would vary across scenarios, and specifically that "sometimes you and Agent Z will know which direction is north, sometimes you won't." Participants were told that there might be alternate ways to describe the correct action (for Agent Z), so to "choose whatever way you think would be clearest, based on the situation and the information available, on a case-by-case basis."

Each scenario page consisted of a paragraph describing the situation, a map depicting it, a question from Agent Z, and a single line for the participant's response. The opening descriptions were written with the goal of entertaining participants in order to keep them engaged with the task; for example, "Unscrupulous nuclear plant operators have secretly stored radioactive waste in a container in a warehouse at Fred's U-Store-It. Fortunately, your Geiger counter has picked out the container, and Agent Z has arrived in her radiation-proof clothing." Descriptions were originally written in English, and translated into Japanese for use by the Kanazawa participants.

Descriptions also gave a brief key to the diagram below them (e.g., "the squares are the containers, the arrow points to the radioactive one"), and in some cases an indication of the compass direction corresponding to the top of the diagram.

Copresent and *remote* scenarios were used. In copresent scenarios, participants were asked to imagine being in a situation along with Agent Z, and icons for the participant and for Agent Z appeared in the accompanying diagrams. For remote scenarios, participants were asked to imagine that they were in a control room looking at a map of an area towards which Agent Z was travelling, and that they were communicating with Agent Z remotely.

In all scenario maps, the two containers were shown as identical filled squares. An arrow pointed to the one holding the target. If the scenario included a landmark, it was shown as an unfilled circle. For copresent scenarios, maps were drawn from the participant's point of view (the participant's icon at the bottom center of the map, oriented upwards).

Below the map, the question from Agent Z always began "Agent Z looks at you and signals:" followed by a question such as "Where's the waste?" or "Which couch has the money?" that again mentioned the target. Above the response line was the tag, "You signal back to him:" or "You signal back to her:".

Design. Three within-participant variables were crossed in a 2 x 2 x 5 design to produce 20 copresent scenario maps. Two other within-participant variables were crossed in a 2 x 2 design to yield four remote scenario maps. In addition, four more remote scenario maps were produced as filler items. Maps were paired with descriptions to form the 28 scenario pages in the questionnaire.

For the 20 copresent scenarios, the first of the within-participant variables was the position and orientation of Agent Z, either facing or perpendicular to the participant's Agent U. The second was the arrangement of the two containers (filled squares) with respect to the participant, either aligned with or perpendicular to the direction the participant was facing. The third, 5-level variable coded the referential resources available in the scenario: baseline (the two agents and the two containers only), landmark/equidistant (baseline plus a landmark equidistant to the containers), landmark/skewed (baseline plus a landmark nearer to one of the containers), compass/north-up (baseline plus inclusion in the description that north is towards the top of the diagram), and compass/north-side (baseline plus inclusion in the description that east or west is towards the top of the diagram—this non-canonical direction of north was predicted to make compass perspectives less attractive). Table 1A shows schematic scenario maps of the 20 resulting conditions.

The four remote scenarios included both landmark and compass information. The location of the landmark (equidistant or skewed) was crossed with the direction of north (up or sideways). Table 1B shows schematic scenario maps.

Due to limitations in the number of participants that could be run, we chose not to counter-balance several variables in the present design. These include which container was designated the target, which side (left or right) a perpendicular addressee was placed, which container was put near a skewed landmark, and which side (left or right) corresponded to north in compass/north-side conditions.

For the US sample, two between-participant variables were crossed: the gender of the participant and the ordering of scenarios in the questionnaire (one of four constructed). A constrained-random ordering was constructed by randomly ordering the 4 baseline copresent scenarios (plus four filler remote scenarios) and then appending a random ordering of the remaining 16 non-baseline copresent scenarios and 4 remote scenarios. (For the Japanese sample, the ordering of scenarios in the questionnaire were fixed.)

Procedure. Participants were run in groups. They were given 30 minutes to complete the questionnaire.

Data Analysis

Following convention, we have called the person giving the communication the "speaker" and the intended recipient the "addressee," even though in our tasks the speaker actually writes instead of speaking and the addressee is hypothetical.

Each written response was analyzed as a description of the target container’s location, followed possibly by one or more redundant re-descriptions. Each description was coded as taking the speaker’s, the addressee’s, a landmark, or a compass direction perspective, or some combination of these. Table 2 shows some examples of each. Each scenario was coded by a single native speaker of the relevant language, supervised and reviewed by the first author.

Unless indicated otherwise, re-descriptions were excluded in the analyses. For example, if a participant wrote “it’s the one to my left and your right,” this would be counted as a speaker-perspective description, with the listener-perspective re-description ignored. This was done to simplify the analysis and make it more intelligible. We felt that it was important to distinguish a participant saying “It’s the one to the south, near me” (a compass perspective description, followed by a redundant speaker-perspective one) from “It’s the one just north of me” (a single hybrid compass-and-speaker perspective). Lumping together all perspective elements in a string of descriptions would lose this distinction. Furthermore, we thought that re-descriptions were generally a sign of a participant being especially helpful, by providing rephrased redundancy, rather than a symptom of a fundamentally different cognitive representation of the situation from that of a participant providing a single description.

Some participants produced procedural descriptions, giving instructions to their addressee about how to get to the target. As much as possible, these were analyzed using the same perspective categories as used for static descriptions. When the instructions began with the addressee at the location indicated in the situation diagram, they were coded as involving the addressee perspective. The instructions might then mention going to (or not past) a landmark, in which case the category would change to addressee & landmark. Sometimes instructions used a generic “you,” locating “you” somewhere in the scenario (e.g., the first L&C example in Table 2); such instructions were judged not literally to involve the addressee’s perspective, unless the generic “you” was initially located at the addressee’s location in the scenario diagram. Examples of procedural instructions and how they were coded are included in the S&L and L&C entries in Table 2.

Results and Discussion

44% of the participants produced at least one re-description, and on average participants produced 26.2 descriptions in total for the 24 scenarios. A two-way analysis of variance was performed with Culture (US or Japanese) and Gender as factors. US participants were more likely to produce a re-description than Japanese participants (59% US vs. 29% Japanese, $F(1,135) = 14.9$, $p < .001$). A main effect was also found for Gender, with females more likely to re-describe than males (51% of females vs. 36% of males, $F(1,135) = 4.02$, $p < .05$). Culture and Gender interacted ($F(1,135) = 4.02$, $p < .05$) in that the effect of gender was due entirely to the US sample, in which 75% of females but only 43% of males produced one or more re-descriptions.

Table 3 shows the overall frequency with which participants used various perspective elements in their initial descriptions. (Note that each element occurs in multiple perspective categories, as outlined in Table 2; for example, landmarks are involved in the L, S&L, A&L, and L&C perspective categories.) When scenarios specified the addressee’s location, participants made use of this information most of the time (76% US, 70% Japanese). Compass direction and landmark information, when supplied, were used less often, but not infrequently. However, speaker information was used in only about one-tenth of the scenarios in which it could have been. Note the striking similarity between the ordering and relative frequency of use of perspective elements across the US and Japanese samples.

A two-way ANOVA with Culture and Gender factors was performed on each of the four measures shown in Table 3. The only reliable differences found were main effects of Culture for the addressee and landmark measures. Though significant, the size of these effects were small especially in comparison to the overall differences in frequency of use of the various perspective elements. On average, US participants referenced the addressee in their initial descriptions 76% of the time, compared to 70% for Japanese participants ($F(1,138) = 4.30$, $p < .05$). This difference is opposite to what one would predict according to a cultural stereotype of Japanese being more deferential or polite than Americans. On the other hand, Japanese participants used landmarks in 38% of the scenarios in which they were available, compared to 28% for US participants ($F(1,138) = 6.70$, $p < .05$). A possible explanation lies in (anecdotally reported) differences in how US and Japanese urban environments are marked. A relative lack of street signs in Japanese cities may require more frequent reference to landmarks in direction giving, and so a general tendency to communicate in landmark terms. Considering that in many scenarios participants could choose to use landmark perspectives instead of addressee perspectives, an overall Japanese preference for landmark perspectives might also underlie the prior finding that Japanese participants were less likely to use addressee perspectives.

Table 4A shows for each perspective the proportion of participants who used it in their initial description, averaged over the four scenarios in each of the five types of copresent scenarios. As before, the overall pattern was similar across cultures. For both the US and Japanese samples, in each case, a majority of participants used their addressee's perspective. At least one-fourth used neutral perspectives (based on landmarks or compass directions) when available, except for the landmark/equidistant scenarios (in which the landmark's location did not easily distinguish between the two containers). Differences in the direction of north (distinguishing between the two compass conditions) had no apparent effect on use of compass directions. In no scenario type was the speaker's perspective used by more than 12% of participants. Overall, Schober's (1993) finding that in non-interactive situations most speakers use their addressee's perspective was replicated.

For each scenario type in Table 4A, a Chi-square test was performed to investigate differences between the US and Japanese samples with respect to their perspective choices. (For these analyses, "Other" perspectives were disregarded.) Only for the landmark scenarios were reliable differences found ($\chi^2(4) = 32.88$, $p < .001$ for landmark/equidistant; $\chi^2(4) = 29.26$, $p < .001$ for landmark/skewed). For the landmark/skewed scenarios, US participants favored mixed addressee-and-landmark relative to pure landmark perspectives, whereas Japanese participants were more divided. For the landmark/equidistant scenarios, US participants were less likely relative to Japanese participants to use either mixed or pure landmark perspectives. Again, a stronger preference to talk in terms of landmarks for Japanese would be consistent with these results.

For the four remote scenarios, in which participants imagined communicating from a remote location to Agent Z prior to Z's arrival at the scene, the trend was for compass perspectives to be more often used in initial descriptions than either landmark or mixed compass-and-landmark perspectives (see Table 4B). However, the US and Japanese samples differed in their perspective choices (disregarding the Other responses, $\chi^2(2) = 36.55$, $p < .001$). Although about 17% of both samples used mixed perspectives, compass perspectives dominated landmark perspectives for the US sample (65% vs. 17%), whereas the Japanese sample showed little preference between them (43% vs. 39%). Summing across the four remote scenarios, the average US participant used compass perspectives 2.54 times compared to 1.71 times for the average Japanese participant ($t(138) = 3.51$, $p < .001$).

Recall that the four scenarios comprising each scenario type represent four different spatial configurations of speaker, addressee, and containers that result from crossing two factors: the relationship of addressee to speaker (facing or perpendicular), and the relationship of the two containers to the speaker (near/far or left/right). Participants' choice of perspective was influenced by these spatial characteristics of the scenarios. For example, Table 5 shows for each of the 20 types of copresent scenarios the percentage of participants who referenced the speaker's perspective in their initial description of the location of the target. In each sample, for most of the scenario types (columns in Table 5) one of the scenarios yielded conspicuously more speaker references than the other three. However, which configuration was thus associated with speaker references differed across the scenario types: configuration PN for the baseline and compass/north-side scenarios, but configuration FN for both landmark and the compass/north-up scenario types. (We will call these five the "speaker-associated" scenarios.)

The configurations had been designed to test a hypothesis that participants would tend to avoid perspectives requiring a left/right discrimination, and thus it had been expected that in all five scenario types configuration PN would yield more speaker perspectives, relative to the other three configurations. However, this was true for only two of the five scenario types. A near/far discrimination from the speaker's perspective, contrasted to a left/right one from the addressee's, was not itself sufficient to lure participants to adopting a speaker perspective.

Upon closer inspection of Table 5 it can be seen that what ties together the five speaker-associated scenarios is that only in these cases was the target located in the container closest to the speaker. This pattern of responses would be consistent with participants having two sometimes conflicting desires: to adopt their addressee's perspective in addressing them, and to choose a perspective allowing use of the relation "near" (or related relations such as "in front of", "closest to," etc.). This qualitative pattern generally holds for both the US and the Japanese samples, though there are some variations in the distribution of speaker perspectives that are not accounted for by this hypothesis (such as the 16% of the Japanese sample who referenced the speaker's perspective in baseline scenario PL).

Discussion. There are four main findings of this first experiment, in which participants wrote brief messages specifying the location of one of two identical objects for a person with a different perspective on the scene. First, we replicated Schober (1993) in finding that participants usually take the perspective of their addressees (as was summarized in Table 3), and found this even when neutral perspectives were available.

Second, we found that choice of perspective can be sensitive to configural aspects of the situation in which the speaker and addressee are embedded. Perspective choices varied considerably and fairly consistently over different scenarios within each category. However, we found little support for our initial hypothesis that advantages in the discriminability of near/far relationships over left/right relationships were the configural aspects that would influence participants' choice of perspective. Instead, the distance between potential points of reference and the target object appeared to be a critical aspect, sometimes counteracting the general trend to take the addressee's perspective. The perspective of whoever was closer to the target item apparently became more attractive by virtue of this proximity.

The third main finding is a null result, and so should be considered cautiously. In the participant population we studied, males were no more likely than females to use compass directions, and females no more likely than males to use landmarks in their spatial descriptions contrary to evidence from differences in men and women's navigational styles. The only gender difference found was that female participants tended to describe the situation more thoroughly, following their initial description with a re-description from a different perspective; however, for unknown reasons this was true only for the US sample.

Finally, the results originally obtained at Stanford University were replicated to a remarkable degree at Kanazawa University, despite large linguistic and cultural differences. Only two systematic differences were found, in terms of landmark use and verbosity of responses. The Japanese participants tended to use landmark information more often, perhaps due to differences between American and Japanese cities. They also tended to be more concise in what they wrote, producing far fewer re-descriptions than the US participants did, and (perhaps consequently) showing no gender differences in this regard, unlike the US sample. A null result worth emphasizing is that despite cultural stereotypes suggesting a relative Japanese emphasis on politeness, the two samples did not differ in their overall use of the addressee's perspective.

Experiment 2: Tell Self

In Experiment 1, the task was to explain to another the location of a target item. We hypothesized that participants were expending mental effort to find a descriptive strategy that their addressee would easily understand. Experiment 2 tested this hypothesis by providing a condition in which participants were not communicating to another, but describing the scene for themselves. We expected participants to make use of the other person's perspective far less often than in Experiment 1, preferring instead to describe scenes from their own point of view or in a perspective-independent, neutral way. Indeed, we expected participants to make use of the others' perspectives only when doing so would allow for a simpler representation.

Method

Participants. 24 undergraduates (12 males, 12 females) at Stanford University and 54 undergraduates (26 males, 28 females) at Kanazawa University from the same participant pools used in Experiment 1 participated.

Materials. A variant of the questionnaire used in Experiment 1 was prepared. The cover story was changed to explain that they, Agent U, would be undergoing pre-mission briefings in which they would be shown where they, Agent Z, and the relevant objects would be located in upcoming actual missions. Their task was to make a brief note to themselves of the location of the hidden target; they would bring this note with them on the mission to help them remember. Individual scenario descriptions were slightly modified to place verbs in the future tense. Scenario maps were identical to those used in Experiment 1. Below these maps, participant's responses were prompted with a statement like "The Encoder Pad asks: Where's the waste?" followed by the tag "You write for yourself:".

Design. The design was identical to that used in Experiment 1, with the exception that since Agent Z's gender was never specified this was not used as a between-participants variable.

Procedure. As in Experiment 1, participants were run in groups and given up to 30 minutes to complete the questionnaire.

Data Analysis

The same coding scheme was used as in Experiment 1 (refer to Table 2), with the appropriate changes to take into account that the participant was addressing him- or herself, not Agent Z. It was still possible for the participant to take Agent Z's perspective, so code A should now be taken to stand for Agent Z, and instances of *addressee* in Table 2 replaced with *Agent Z.*

Results and Discussion

Overall Results. 76% of the participants produced at least one re-description, and on average participants produced 32.5 descriptions in total for the 24 scenarios. An ANOVA with Culture and Gender as factors found no significant main effects nor interactions.

Table 6 shows the overall frequency with which participants used various perspective elements in their initial descriptions. In terms of the use of speaker vs. addressee/Agent Z perspectives, the results are opposite those found in Experiment 1 (Table 3). For the 20 copresent scenarios, the speaker's perspective was used 50% of the time, the addressee's perspective only 18% of the time. Indeed, both landmark and compass information was used more often than addressee information, at levels roughly comparable to those found in Experiment 1. As before, the overall pattern was similar for US and Japanese participants.

A two-way ANOVA with Culture and Gender factors was performed on each of the four measures shown in Table 6. Only one marginal effect was found, a Culture X Gender interaction in the use of compass perspectives ($F(1,74) = 3.82$, $p < .06$). Although US men and Japanese men used compass perspectives at the same rate (62% of the time for US men, 64% of the time for Japanese men), US women used compass directions more often than did Japanese women (76% of the time for US women, 53% of the time for Japanese women). This interaction was not predicted and is unexplained. The differences found in Experiment 1 between the US and Japanese samples with respect to overall addressee and landmark use were not replicated, suggesting these cultural differences may be highly task-specific.

Table 7A shows the proportion of participants using each perspective type, averaged over each scenario type. For baseline and landmark/equidistant scenarios, in which no neutral perspectives were easily available, speaker perspectives were favored by a majority of both US and Japanese participants. However, for the other types of copresent scenarios more amenable to neutral description, neutral perspectives were usually most common. Again, this pattern of results markedly differs from that found in Experiment 1 (see Table 4A), in which addressee perspectives were favored for all copresent scenario types.

For each scenario type in Table 7A, a Chi-square test was performed to investigate differences between the US and Japanese samples. (For these analyses, "Other" perspectives were disregarded.) The only reliable difference found was for landmark/equidistant scenarios ($\chi^2(4) = 14.9$, $p < .01$); as can be seen in Table 7A, compared to Japanese participants, US participants were more likely to use Agent Z perspectives and less likely to use landmark perspectives in this case. Note that to use a landmark perspective in the landmark/equidistant case, English participants would have to write something like "looking towards the landmark, it's the one on the left." Japanese participants appeared to use shorter constructions, such as "the left one toward the landmark," perhaps another manifestation of a greater Japanese familiarity with describing scenes in terms of landmarks.

Table 7B presents the results for the remote scenarios. Compared to Experiment 1 (see Table 4B), the Japanese participants in both experiments responded almost identically: split roughly evenly between compass and landmark perspectives, with about 20% using a mixed compass and landmark perspective. Unlike Experiment 1, the US and Japanese participants in Experiment 2 did not differ significantly in their perspective choices according to a Chi-square test (disregarding the very rare "Other" responses).

As in Experiment 1, use of the speaker's vs. addressee's perspective varied according to configural aspects of the scenarios (see Table 8). Again, there is some evidence of a bias towards perspectives affording use of the relation *near*, but the pattern is not so straightforward. In Experiment 1 (see Table 4), evidence for a *near* bias could be seen in increased use of the generally non-preferred speaker perspective when the target was in the container closer to the speaker. In the present experiment, it can be seen in increased use of the generally non-preferred addressee perspective when the target was in the container closer to the addressee: configuration FN for the baseline and compass/north-side scenarios, and (for the US sample only) configuration PL for the landmark perspectives.

Discussion. This experiment, in which participants wrote brief messages specifying the location of one of two identical objects for themselves, replicated some aspects of the first experiment. However, it also demonstrated large differences between participants' choice of perspectives when making notes to themselves and their choices when communicating to another. Participants were like those in Experiment 1 in exhibiting proximity effects (often tending to choose the perspective of whoever was closest to the target). They differed in generally preferring neutral perspectives, except when these were not available (in the baseline condition) or difficult to use (in the landmark/equidistant condition), in which case they generally preferred to use their own character's perspective (as shown in Table 7A).

Participants' descriptions also made use of Agent Z, often when describing the scene from Z's point of view affording use of the relation *near*. When participants used Agent Z's perspective in describing a scene to themselves, they were most likely considering Agent Z to be nothing more than a convenient landmark. The frequency with which this happened raises the possibility that in Experiment 1 some use of the addressee's perspective may have also have had this character, even though participants were communicating to another.

As in Experiment 1, the US and Japanese samples behaved remarkably alike. The main difference found involved responses to the landmark/equidistant scenarios, in which Japanese participants were more likely to make use of landmark perspectives. Although this was consistent with patterns in Experiment 1 of increased landmark use by the Japanese participants, the larger trend found in Experiment 1 that landmarks were mentioned more often by Japanese than US participants was not apparent in Experiment 2.

Experiment 3: Ask Other

In this experiment, participants were asked to imagine being in scenarios much like those in Experiment 1, in which a target was hidden in one of two identical objects. But rather than being told (by an arrow on the map) which of the two was the target, participants had to obtain this information from their imagined partner, Agent Z. They were asked to produce a question that could be answered with a simple yes or no. Thus, participants had to pick out by means of a locative expression one of the two potential targets and ask Agent Z if it did (or did not) contain the target. Either a yes or a no response would unambiguously specify the target, since there were only two to choose from.

If politeness were the only factor, then participants should frame questions using Agent Z’s perspective, and arbitrarily choose between the two objects in selecting one to mention in their query. But if communicative balance and ease are the primary determinants of question framing, choice of whose perspective to use and which object to select would be interdependent. If, as in the previous experiment, participants were continuing to express a preference for the relation *near* over *far*, a perspective and an object would be picked between which this relation held. For example, participants would prefer to ask “Is the bomb in the suitcase near you?” rather than “Is the bomb in the suitcase to my left?” or “Is the bomb in the suitcase further from you?”

Method

Participants. 64 Stanford undergraduates (33 men, 31 women) and 54 Kanazawa undergraduates (28 men, 26 women) participated in order to fulfill course requirements or for a small cash stipend.

Materials. A variant of the questionnaire used in Experiment 1 was prepared. The cover story was changed to explain that Agent Z would be helping them to accomplish various missions, and that their Encoder Pads would only allow a simple question to be transmitted to Z and a yes or no response returned. Scenario descriptions were modified to be consistent with the participant’s character having the means but not the information to solve the problem at hand, whereas the converse was true for Agent Z. Participants were prompted for their responses by the tag, “You signal to Agent Z, asking him:” Below each response line participants were told, “Remember, Agent Z can only reply ‘YES’ or ‘NO.’”

Design. A total of 28 scenarios (in addition to an initial sample scenario) were constructed. These divide into three groups: 8 scenarios in which Agents U and Z but no landmarks were present; 16 scenarios in which Agents U and Z were copresent, and in which a landmark was near one of the two containers; and 4 scenarios in which neither Agent U nor Z were present yet in the situation. (No landmark/equidistant scenarios were used in Experiment 3.)

The 8 no-landmark scenarios conform to a 2 x 2 x 2 design, with the factors being: the orientation of Agent Z (either facing U, or perpendicular U on U’s left), the orientation of the containers (aligned with or perpendicular to the direction Agent U was facing), and the presence or absence of compass direction information (absent for baseline scenarios, present for compass scenarios). When compass directions were given, north was always to Agent U’s left or right, though this was not counterbalanced with the other factors.

The 16 copresent/landmark scenarios conform to a 2 x 2 x 2 x 2 design. The first three factors are the same as those used for the no-landmark scenarios. The fourth factor specifies which of the two containers was near the landmark. This factor allows landmark location to be counter-balanced across scenarios, overcoming a limitation of the design of the landmark/skewed scenarios used in Experiments 1 and 2.

Finally, the four remote scenarios conform to a 2 x 2 design. The first factor specifies the orientation of the containers (aligned with or perpendicular to the side of the page). The second specifies which of the two containers was near the landmark, as in the copresent/landmark scenarios. The direction of north was supplied for all four remote scenarios, and was either to the left or the right of the page.

As in Experiment 1, a 2 x 4 between-participant factorial design crossed the gender of participant and the ordering of scenarios in the questionnaire (one of four constructed). (The replication at Kanazawa University used only one ordering of scenarios.) Unlike Experiment 1, the baseline scenarios were intermixed with the others, rather than appearing among the first eight.

Procedure. As in Experiment 1, participants were run in groups and given up to 30 minutes to complete the questionnaire.

Data Analysis

The same coding scheme was used as in Experiment 1 (refer to Table 2).

Results and Discussion

Overall Results. 24% of the participants produced at least one re-description, and on average produced 28.9 descriptions in total for the 28 scenarios. (Analysis of variance found no effects of Culture nor Gender.) These figures are much lower than in the previous experiments. It may be that it is more awkward to ask questions containing redundant descriptions than it is to produce an assertion with redundancies.

Table 9 shows the overall frequency with which participants used various perspective elements in their initial descriptions. Both persons (speaker and Agent Z) were referred to proportionately less often than either landmark or compass information, unlike the pattern in either of the previous two experiments.

A two-way ANOVA with Culture and Gender factors was performed on each of the four measures shown in Table 9. A main effect of Culture was found only for use of addressee perspectives: whereas US participants made use of the addressee's perspective in 37% of the scenarios in which they could, Japanese participants did for only 25% of such scenarios ($F(1,114) = 9.59$, $p < .01$). Furthermore, a Culture X Gender interaction was found for the use of compass perspectives ($F(1,114) = 4.23$, $p < .05$): whereas in the Japanese sample men and women used compass directions about equally often (60% for men, 57% for women), in the US sample men used them less often than women (39% for men, 59% for women).

Table 10 shows the proportion of participants using each perspective type, averaged over each block of scenarios. Overall, for both samples, neutral (compass and landmark) perspectives were most popular, addressee perspectives second most, and speaker perspectives least. Unlike previous experiments, mixed personal/neutral perspectives were rarely used.

For each scenario type in Table 10, a Chi-square test was performed to investigate differences between the US and Japanese samples. (For these analyses the rarely used, "S&N," "A&N," and "Other" perspectives were disregarded.) For the baseline and remote scenario types, the US and Japanese samples did not reliably differ, but there were significant differences between samples for the other three scenario types. In the landmark scenarios, US participants used addressee perspectives more and neutral perspectives less than their Japanese counterparts ($\chi^2(2) = 33.6$, $p < .001$). The pattern also held true for the compass scenarios ($\chi^2(2) = 10.9$, $p < .01$) and the landmark & compass scenarios ($\chi^2(2) = 48.7$, $p < .001$). Apparently, for scenarios involving a choice between personal and neutral perspectives, personal perspectives were relatively more favored by US than by Japanese participants.

To analyze the relationship between perspective choice and spatial relation choice, a subset of the data was selected in which respondents had chosen to use "pure" perspectives, and in which, if they had used a compass perspective, they had used a simple cardinal relation north, south, east, or west. 179 out of 3196 responses (5.6%) were thus discarded. For the remaining responses, Table 11 shows how spatial relation choice was distributed within each perspective type. For any given scenario, spatial relation choice was constrained by the spatial configuration presented to participants and well as by the perspective participants chose to adopt. For example, for some responses in which speaker perspectives were used, the choice of spatial relation was between near and far; for others, it was between left and right; but it was never (for a single scenario) between near and left, or between far and right, etc. However, the data presented in Table 11 are averaged across all scenarios, and allow the relative attractiveness of near vs. left, far vs. right, and all other pairs of relations to be observed.

The pattern of responses shown in Table 11A indicates a clear preference for the spatial relation near over its competitors left, right, and (especially) far. The preference is strongest (and nearly unanimous, 99.0%) for landmark perspectives, which is not surprising given that none of the landmarks used in the scenarios had intrinsic axes allowing left or right to be easily defined. On the other hand, speakers and addressees do have intrinsic axes, and would both equally and relatively easily afford the use of left or right relations. Nevertheless, the preference for near is stronger within speaker perspectives than within addressee perspectives (averaging across samples, 73.5% of speaker responses used near compared to 61.7% of addressee responses). This may be attributed to the status of the addressee perspective as the default or normal perspective from which to describe a scenario. Departures from this norm to use the speaker perspective instead could be justified in circumstances where the speaker but not the addressee perspective afforded using near.

Table 11B shows the relative frequency with which the different cardinal direction terms were used when participants adopted a simple compass perspective. Averaging across samples, north was used most of the time (53.2%), followed by east (25.2%), west (16.8%), and south (4.8%). This is consistent with participants mapping cardinal directions on to the axes of the human body as when reading a map in canonical orientation, with north up, south down, east to the right, and west to the left. Such a mapping would predict the relatively large difference between rates of using north vs. south (i.e., up vs. down) compared to east vs. west (i.e., right vs. left), since the up/down axis is more differentiated in human bodies than left/right and thus simpler to process (see Franklin & Tversky, 1990 and Bryant & Tversky, 1992).

For each of the four perspectives considered in Table 11, a Chi-square test was performed to compare the US and Japanese samples. The two samples were significantly different in their spatial relation choices for all but the landmark perspectives (where the choice of near was nearly unanimous). For both the speaker perspective responses ($\chi^2(3) = 15.1$, $p < .01$) and the addressee perspective responses ($\chi^2(3) = 33.4$, $p < .001$), Japanese participants exhibited a strong preference for right over left, whereas US participants exhibited either the opposite trend (for speaker perspectives) or indifference (for addressee perspectives). We do not know what would account for this pattern of results. The samples also differed with respect to compass perspective responses: compared to US participants, Japanese participants tended to more strongly prefer north to south, and east to west ($\chi^2(3) = 25.9$, $p < .001$). Again, these results are puzzling. The stronger preference for north over south by Japanese informants may be related to a higher cultural significance of north reflected, for example, in the rigid ways capital cities were laid out, with the emperor's palace to the north. The stronger preference for east over west may be related to the previously noted preference for right over left, as east is to the right when one faces north (or holds a map canonically).

Discussion. As in the “tell other” condition, participants in this experiment were communicating with another person, and so it was in their interest to make their messages easily understood by their addressees. On the other hand, as in the “tell self” condition, participants would themselves be using the spatial information, and so it was also in their interest to design their questions to yield answers they themselves could easily understand. Perhaps as a result of this conflict, the overall use of neutral (compass and/or landmark) perspectives was greater than in the prior experiments. Those participants choosing not to use a neutral perspective tended as in Experiment 1 to favor the perspective of their addressee. Apparently in these cases the need to make oneself understood to another took precedence over the need to make oneself understood to oneself.

As in previous experiments, participants were influenced considerably by the geometry of the situation, being biased in favor of perspectives from which the target object was near or (equivalently, in this experiment) in front. Strong interactions between perspective chosen and term of reference used within that perspective further supported participants’ bias towards the relation *near*.

Summary and Comparison of Results from Experiments 1 to 3

At this point, it is worthwhile to examine some general trends across the three situations under which spatial descriptions were produced: tell other, tell self, and ask other. The three fall at different places on a continuum of relative cognitive burden in communication, with more of the burden on the addressee for “tell other,” all of the burden on the speaker for “tell self,” and the burden more evenly distributed between participant and addressee for “ask other.” The expectation is that the relative use of speaker and addressee viewpoints will follow the relative cognitive burden in communication. Figure 3 summarizes the results from the four baseline scenarios in each experiment. In these scenarios, participants had a simple choice between using the speaker's perspective or adopting Agent Z's, since neither compass directions or landmarks were available. For each participant, the proportion of responses to baseline scenarios that used Agent Z's perspective was calculated: Figure 3 plots the means of these proportions. The proportion taking Agent Z's perspective diminished dramatically as the cognitive burden on Agent Z diminished. Note, however, that even in the "tell self" condition, Agent Z's perspective continued to be used. In these cases, the participant seemed to be using Agent Z as a landmark, especially to avoid using *left* or *right*. The US and Japanese samples hardly differed on these measures.

The three situations were also expected to affect the relative use of personal versus neutral perspectives, that is landmark or compass directions. Using a neutral perspective avoids having to decide whose personal perspective to use, speaker's or addressee's. Thus, use of a neutral perspective should be greater in the "ask other," where the communication is more equal, than the "tell other" situation. Communicative balance is not relevant in the "tell self" situation. In that case, participants are writing a memory aid for a future situation, and so may be attracted more to neutral perspectives since they remain valid even if the speaker's and Agent Z's locations should change between when the note is written and when it is used.

Figure 4 summarizes results from the scenarios in each experiment in which participants had a choice between using a personal perspective and using either a compass-based or a (skewed) landmark-based neutral "perspective." (Scenarios in the "ask other" experiment (Experiment 3) in which both landmark and compass information was available were excluded.) Mixed perspectives, such as "the one southwest of me," were coded as personal in this analysis. For each participant, the proportion of these scenarios using a pure neutral "perspective" was calculated; Figure 4 plots the means of these proportions. As expected, neutral as opposed to personal perspectives were more popular in the "ask other" and "tell self" situations. Neutral perspectives were used about half the time in the "ask other" and "tell self" situations, but only about a quarter of the time in the "tell other" situation, where use of personal perspectives dominated.

Figure 4 also shows a complex pattern of differences between the US and Japanese samples across experiments. In terms of landmark use, in each experiment Japanese participants used them more than did US participants, and this may be due to increased reliance on landmarks in everyday navigation in Japan. However, the pattern for compass-based perspectives is inconsistent and difficult to explain – no difference for the "tell other" experiment, larger Japanese use for "ask other," and larger US use for "tell self."

Experiment 4: Testing the Near Bias

The results of the previous experiments suggested that people prefer locating a target using *near* to locating a target using *far* or the direction terms *left* and *right*. However, the alternative descriptors were not varied systematically in the previous experiments. Experiment 4 was designed to investigate the near bias systematically. To this end, a short questionnaire was administered to a large number of participants. The questionnaire presented a single scenario of the type used in Experiment 1 ("tell other"), and gave the participant a two-alternative forced choice between a speaker- and an addressee-based description. The geometry of the situation was varied across conditions, so that in making this choice some participants were comparing a description using *near* with one using *far*, other participants a description using *near* with one using *left*, etc. A simple mathematical model was developed to try to predict the proportion of participants choosing a description, given only the pair relational terms that were competing in the forced-choice.

Method

Participants. 209 Stanford undergraduates (86 men, 123 women) participated in order to fulfill course requirements.

Materials. A variant of the questionnaire used in Experiment 1 was prepared. A single scenario was used:

Imagine, if you will, that you have hidden a present for your friend [X] in one of two identical boxes in your living room. The situation is shown below. You and [X] are both facing the center of the room (between the two boxes). The arrow points to the box with the present.

(For half the forms, [X] was the male name *Dirk*, for the others the female name *Gwen*.) Following this description was a diagram similar to that shown in Figure 2. One of eight different diagrams was used, based on a 2 x 2 x 2 design: the addressee (Dirk or Gwen) was depicted either at the top of the diagram (across from the participant) or at the left of the diagram (diagonally in front and to the left of the participant); the boxes were either aligned with or perpendicular to the direction the participant was facing; and either the top (or left) or bottom (or right) box was marked as containing the target.

The form then asked: “Without gesturing, how would you tell [X] where [his or her] present is? Please consider each of the options given below, and circle the one that is most similar to what you would actually say to [him or her].”

Option 1 was a speaker-centered description and option 2 was an addressee-centered one; for example, “It’s in the box to my left” and “It’s in the box nearest to you.” The 2 x 2 x 2 design used entailed that each participant chose between the spatial relations in one of the pairings given in Table 12.

Procedure. One-page experiment forms were prepared and inserted into booklets of approximately 30 different questionnaires from various studies. Booklets were distributed to participants to fill out at their own pace.

Results

There were no effects of gender found, either of the actual participant or the imagined addressee (Dirk vs. Gwen), nor was there a significant interaction.

Table 13 shows the proportion of participants choosing the description that took the addressee’s point of view as a function of the spatial relations in alternative descriptions. The near bias is again apparent in the first row and the first column in the table. Almost all (96% and

93%) of participants took the addressee's perspective when the addressee was near the target, while few (26% and 10%) took the addressee's perspective when the speaker was near the target. In choosing between the relations *left* and *right*, participants are indifferent (50% and 44% choosing *left*). In choices between *left* or *right* and *far*, *far* fared well — 70% chose the addressee's perspective when the target was far from the addressee (and to the right of the speaker), but only 36% chose the addressee's perspective when the target was far from the speaker (and to the left of the addressee).

A simple model was constructed to account for the results as follows. Each spatial relation (*near*, *far*, *left*, and *right*) was assumed to have some level of inherent attractiveness. The proportion of participants choosing the description from the addressee's perspective was predicted to be the ratio of the attractiveness of the spatial relation in that description to the sum of the attractiveness ratings of the two spatial relations competing for the participants' favor. In mathematical notation:

$$P_{\text{addressee}} = A(R_{\text{target,addressee}}) / [A(R_{\text{target,addressee}}) + A(R_{\text{target,speaker}})]$$

where $P_{\text{addressee}}$ is the proportion of participants choosing the addressee's perspective, $R_{\text{x,y}}$ is the spatial relationship of x with respect to y, and A(r) is the attractiveness of spatial relationship r.

Constraining A(*left*) = A(*right*), this model has three parameters. A good fit (with a mean absolute error of 4.64%) was obtained with A(*left*) = A(*right*) = 1.0, A(*far*) = 2.4, and A(*near*) = 22.7 (see Figure 5). In other words, in the framework of this simple model, *far* was about twice as attractive as *left* or *right*, and *near* was an order of magnitude more attractive than *far*. Note that this fit was obtained without predicting any overall bias towards taking the addressee's perspective. (The reason for noticeably aberrant case — in which 26% of participants preferred "to your right" to "nearest to me," whereas the model predicted 4% — is unknown.)

Discussion. This experiment provides further evidence for a strong bias towards choosing to describe spatial relationships with the relation *near*, and in a crude attempt to quantify the strength of this bias found a tendency to use *near* of an order of magnitude larger than *far*, *left*, or *right*. Furthermore, it demonstrates that discriminability may be also play a role, in so far as both the highly discriminable terms *near* and *far* are preferred (by at least a factor of 2) to the more difficult to discriminate *left* and *right*. Puzzlingly, unlike in Experiment 1, a bias towards taking the addressee's perspective was not observed, perhaps because participants were not immersed in the task, which was very brief and embedded in a long questionnaire of unrelated tasks.

General Discussion

How do people describe simple spatial situations, here, the location of one of two identical objects? This question was investigated in simple schematic scenes. Such descriptions require a reference or origin object and terms of reference. While disagreeing on many issues, linguists and psycholinguists have agreed that the primary origins are a person or an object, and the primary terms of reference are those depending on a person or viewpoint (*left*, *right*, *front*, and *back*) and those depending on the environment (*north*, *south*, *east*, and *west*). Does this analysis provide insight into the paradigmatic scenes investigated here? The minimal scenes included only the two objects and the viewpoints of the participant and a partner. The enriched scenes also included a landmark or an indication of the cardinal directions. In the minimal scenes, participants' descriptions were limited to their own perspective or that of their partner. In the enriched scenes more typical of everyday spatial settings, participants could use those perspectives or adopt perspectives based on landmarks or cardinal directions. Participants described the location of the target object in one of three tasks that varied the relative cognitive burden of speakers and addressees. In the "tell other" task, participants knew the target object and sent a brief message conveying that information to their imaginary partners. In the "ask other" task, their partners knew the target object, and participants asked their partners a yes/no question whose answer would reveal the target object to them. In the "tell self" task, participants recorded a brief description for themselves, to be used at a later time to identify the target object. Participants were college students in the US and Japan.

Several interdependent factors, social, spatial, and cognitive, might affect perspective choice. One is politeness. It is polite to take into consideration the viewpoint of an addressee, both literally and figuratively. In work that inspired the present experiments, Schober (1993) investigated the minimal “tell other” condition, finding that in most cases, participants took the perspective of the addressee. Another is the difficulty of describing a spatial scene, a combination of spatial and cognitive factors. For example, descriptions using *left* or *right* are known to be more difficult than those using *front* or *near*. Using *left* and *right* requires discrimination of spatial symmetry whereas using *front*/*behind* or *near*/*far* do not.

For the "tell other" task, use of the addressee's perspective dominated but not for the "ask other" and even less so for the "tell self" tasks. Mirroring these data were those for use of landmarks and cardinal directions, which were lower for the "tell other" task than for the "ask other" and "tell self" tasks. Other salient phenomena in the present data were the use of the partner’s location as a landmark and the bias toward *near*. These findings make it clear that choosing which of several ways to describe a situation is not simply adopting the point of view of one’s addressee, nor selecting whatever point of view is simplest based on the spatial layout of the situation. If it were solely the former, differences in spatial layout would not have yielded the observed differences within each task situation or between different scenarios within a situation. If it were solely the latter, whether self or other was being addressed and whether declarations or questions were being produced would not have yielded differences within each scenario between the task situations.

In fact, there was no evidence for politeness per se. Taking the addressee’s perspective can be fully accounted for by considerations of facilitating joint communication. Although participants overwhelmingly took the addressee’s perspective in the “tell other” task, they did so to a far lesser extent in the “ask other” task, where politeness is no less a factor. Taking the addressee’s perspective in the “tell other” task seems to be more in the interests of increasing the success of the communication than in the interests of being polite. In that situation, the addressee has the harder task, both to understand the message and to use it to determine the critical object. The participant can ease the task of understanding by framing an utterance from the addressee's point of view. In the "ask other" task, both addressee and participant must understand each others' utterances, so the cognitive burden is more equally distributed. In that case, participants were almost as likely to use their own perspective as to use their addressee's. The replication of all of the effects in the data collected in Japan strengthen these and our other conclusions. That taking the perspective of addressees is primarily to insure successful communication is also supported by Schober's (1993) finding that speakers were less likely to adopt their addressees' perspectives in an interactive situation than when their addressees were absent. In an interactive situation, speakers know when they have succeeded in communicating and may easily discover that their addressees understand utterances from other perspectives. When addressees receive no feedback from speakers, then speakers go to extra lengths to guarantee that their message is understood.

Using Agent Z's perspective in the "tell self" condition is neither a matter of politeness nor of joint communication as the communication is for self only. These cases are not genuine cases of taking Z's perspective, but rather of using Z as a landmark. Using Z as a landmark increased when Z's position relative to the target made it easy to locate the target, most notably when the target was closer to Z and when using the participant's perspective would have required a left/right discrimination. The frequent use of Z's perspective in the "tell self" task suggests that many cases of using the participant's or the addressee's perspective in the other tasks and in others' studies were cases of using the participant or the addresseee as landmarks, rather than genuine cases of what is usually thought of as perspective taking. There is no simple way of knowing what uses are landmarks and what are perspective-taking.

Use of landmarks and cardinal directions increased under circumstances similar to those that increased use of participant perspective, that is, when the cognitive burdens of speaker and addressee were more equal, when the message was for self, and when a personal perspective would have required a left/right discrimination. Some aspects of the data suggested that of the neutral perspectives, the cardinal directions were preferred to landmarks, though this did not happen consistently. Each of the nonpersonal perspectives has its advantages and disadvantages, in general and in this situation. Landmarks were visible in the diagrams of the scenes just as they are in real scenes. The cardinal directions were not, again, as they are not in real scenes, though they were indicated, unlike in real scenes. However, in real scenes the cardinal directions are unchanging and ever-present whereas landmarks, especially relatively small ones such as those in the present scenarios, are mutable. Thus, although landmarks and cardinal directions were often preferred to personal perspectives, neither dominated the other.

Previous claims notwithstanding (e.g., Bever, 1992), women did not prefer landmarks nor did men prefer cardinal directions. In fact, there were no recurring gender effects at all. Many other studies in spatial cognition have failed to find gender differences (e. g., Franklin & Tversky, 1990; Taylor & Tversky, 1992). These negative findings are rarely if ever cited in the literature on individual differences, perhaps because the focus of these studies was on spatial cognition and not on individual differences. Indeed, many studies in cognition do not report or even test for gender differences. Given the increasing interest in the topic and the increasing claims of differences, researchers in spatial cognition can easily contribute to the body of knowledge about gender and spatial cognition simply by including gender as a factor in analyses and reporting the findings.

Difficulty of discrimination did play a role in formulating spatial descriptions. Nonpersonal perspectives and participant's perspective were more likely to be selected when the alternative was *left* or *right* than when the alternative was *closer* or *near* or *in front of*. These terms are easier to produce and understand as they do not require projections of directions in space. A related phenomenon occurs in languages that use the cardinal directions to locate objects in space rather than personal direction terms. Speakers of those languages will occasionally use *front* and *back* instead of the cardinal directions, but rarely use *left* and *right* instead of the cardinal directions (Pederson, personal communication, 1995).

Difficulty of discrimination is related to the final phenomenon to emerge from this research, the bias to locate objects using *near* or *closer* rather than *far* or *farther*. This cannot be due to language per se, as *far* is the neutral or unmarked member of the pair and is easier to process (Clark & Clark, 1977). Rather, terms like *near* seem to be preferred to their opposites because they are more informative in describing space. When an object is described as near a landmark, there is less uncertainty about its location than when it is described as far from a landmark. An object described as near is within an understood distance from the landmark, depending on the nature of the object and landmark (Morrow & Clark, 1988); however an object described as far from a landmark may lie in any direction and at any distance. Of course, because there were only two alternatives in the present situation, the actual uncertainty was no greater for *far* than for *near*, making it all the more surprising that *near* was so greatly preferred.

Terms like *near* were preferred not just to terms like *far* but also to terms like *left.* This points to yet another advantage to locating objects as near to a landmark or person. *Near* conveys distance but does not specify direction, unlike *left*/*right*, *north*/*south*, and so on. It only requires discriminating an approximate region around a landmark and does not require determining a direction, a sometimes difficult task. Although *near* was used frequently in our tasks, it has not been considered in many treatments of deixis nor in most treatments of perspective. This may be because *near* does not indicate direction, and even with *far*, it does not partition space neatly into four quadrants, like the other sets of relational pairs. In addition, *near* and *far* lack the symmetry of the other relational pairs. *Near* suggests a relatively small region circumscribing a landmark whereas *far* does not suggest a readily understood spatial region.

Taken together, these experiments demonstrate that the context to which people are attuned in producing such descriptions intertwines both social and cognitive elements. Successful communication results from a collaboration between speaker and addressee (Clark & Wilkes-Gibbs, 1986). In order to insure that a message is communicated, the speaker needs to take into account the addressee's perspective, but also needs to take into account the relative cognitive burdens of all participants in both formulating and comprehending utterances. Ease of formulating and comprehending utterances depends in turn on the spatial situation and the various means to describe it.

Describing the location of an object in space requires grounding in a reference object and terms of reference. Traditional categories of reference frames have been grounded in a person or an object and have used terms of reference that depend on a viewpoint or the intrinsic sides of a person or object, or terms of reference that depend on a surrounding environmental. Language use even in simple paradigms does not always fit easily into the elegant categories carved by linguists. Although extended descriptions of various environments show correlations with the categories proposed by Levinson, there are exceptions (e.g., Taylor and Tversky, 1996) and cases not readily incorporated such as whose perspective (Schober, 1993, 1995). The correlations that exist between reference object and terms of reference seem to be a consequence of typical ways of experiencing environments, from a fixed point at the edge, such as an entrance, for the relative frame of reference, from touring an environment for the intrinsic frame of reference, and from a fixed point above for the extrinsic frame of reference. Yet other limitations of the traditional analyses emerged from the present project. In many cases, it was not possible to know whether speakers were using another's perspective or using the other as a landmark. The present encounter of real-world data with elegant systems for categorizing reference frames challenges attempts to pigeon-hole speech. Not only do many common uses not fall into the traditional pragmatic categories, but also many cases of language use are simply ambiguous with respect to the traditional pragmatic categories. Language in practice reflects the subtle interplay of cognitive, situational, and social factors.

Acknowledgements

This research was supported in part by Interval Research Corporation. We thank Herbert Colston for proposing and collecting data for Experiment 3, Michitaka Ohgishi for collecting data at Kanazawa University, and our research assistants, Dana Peterson, Kim Saccio, and David Rochefort.

References

Bever, T. (1992). The logical and extrinsic sources of modularity. In M. A. Gunnar & M. Maratsos (Eds.), *The Minnesota symposia on child psychology: Vol. 25. Modularity and constraints in language and cognition* (pp. 179–212). Hillsdale, NJ: Erlbaum.

Brown, P., & Levinson, S. (1987). *Politeness: Some universals in language use*. Cambridge, UK: Cambridge University Press.

Bryant, D. J., & Tversky, B. (1992). Assessing spatial frameworks with object and direction probes. *Bulletin of the Psychonomic Society*, *30*, 29–32.

Buhler, K. (1982). The deictic field of language and deictic words (pp. 9–30). In R. J. Jarvella & W. Klein (Eds.), *Speech, place and action*. New York: Wiley. (Translation of part of 1934 book in German.)

Clark, H. H. (1973). Space, time, semantics, and the child. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 27–63). New York: Academic Press.

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

Clark, H. H., & Clark, E. V. (1977). *Psychology and language*. San Diego: Harcourt Brace Jovanovich.

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

Ehrich, V., & Koster, C. (1983). Discourse organization and sentence form: The structure of room descriptions in Dutch. *Discourse Processes, 6*, 169-195.

Fillmore, C. (1975). *Santa Cruz Lectures on Deixis*. Bloomington, IN: Indiana University Linguistics Club.

Fillmore, C. (1982). Toward a descriptive framework for spatial deixis. In R. J. Jarvella & W. Klein (Eds.), *Speech, place and action* (pp. 31–59). London: Wiley.

Franklin, N., & Tversky, B. (1990). Searching imagined environments. *Journal of Experimental Psychology: General, 119*, 63–76.

Galea, L. A. M., & Kimura, D. (1992). Sex differences in route-learning. *Personality and Individual Differences, 14*, 53–65.

Herrmann, T. (1989). Partnerbezogene Objektlokalisation—ein neues sprachpsychologisches Forschungsthema. *Bericht Nr. 25, Arbeiten der Forschergruppe "Sprechen und Sprachverstehen im sozialen Kontext," Heidleberg/Mannheim.*

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago.

Levelt, W. J. M. (1984). Some perceptual limitations on talking about space. In A. J. van Doorn, W. A. van der Grind, & J. J. Koenderink (Eds.), *Limits on perception* (pp. 323–358). Utrecht, The Netherlands: VNU Science Press.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Levelt, W. J. M. (1996). Perspective taking and ellipsis in spatial descriptions. In P. Bloom, M. A. Peterson, L. Nadel, & M. Garrett (Eds.), *Space and language* (pp. 77–107). 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, *Space and Language* (pp. 109-169). Cambridge, MA: MIT Press.

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

Morrow, D. G., & Clark, H. H. (1988). Interpreting words in spatial descriptions. *Language and Cognitive Processes, 3*, 275–291.

Pederson, E. (1995). Personal communication.

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.

Talmy, L. (1983). How language structures space. In H. L. Pick, Jr. & L. P. Acredolo (Eds.), *Spatial orientation: Theory, research and application* (pp. 225–282). New York: Plenum Press.

Taylor, H. A., & Tversky, B. (1992). Spatial mental models derived from survey and route descriptions. *Journal of Memory and Language*, *31*, 261–282.

Taylor, H. A., & Tversky, B. (1996). Perspective in spatial descriptions. *Journal of Memory and Language*, *35*, 371–391.

Tversky, B. (1992). Distortions in cognitive maps. *Geoforum*, *23*, 131-138.

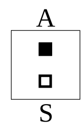
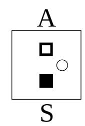
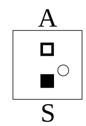
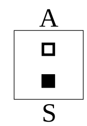
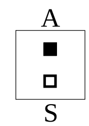
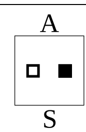
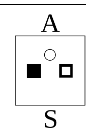
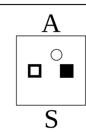
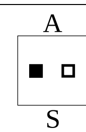
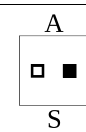
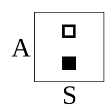
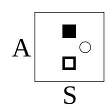
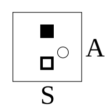
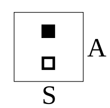
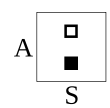
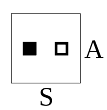
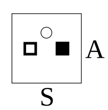
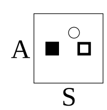
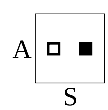
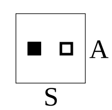
Tversky, B. (1996). Spatial perspective in descriptions. In P. Bloom, M. A. Peterson, L. Nadel, & M. Garrett (Eds.), *Space and language* (pp. 463–491). Cambridge, MA: MIT Press.

Ullman, S. (1996). *High-level vision: Object recognition and visual cognition.* Cambridge, MA: MIT Press.

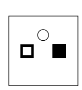
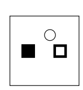
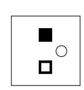
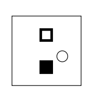
Table 1

Schematic Diagrams of the Scenarios

A. The 20 Copresent Scenarios

Relationship of			Scenario type				
Addressee to speaker	Containers to speaker	Config. code	Baseline	Landmark/ equidistant	Landmark/ skewed	Compass/ north-up	Compass/ north-side
Facing	Near/far	FN					
Facing	Left/right	FL					
Perpen-dicular	Near/far	PN					
Perpen-dicular	Left/right	PL					

B. The Four Remote Scenarios

Direction of north:	Up		Sideways	
Location of landmark:	Equidistant	Skewed	Equidistant	Skewed
				

Note. A and S represent icons for the addressee (Agent Z) and the participant (Agent U), respectively; the squares represent containers, the target filled and the non-target open; and the circles represent landmarks.

Table 2

Perspectives Codes and Examples

Code	*Basis or Bases*	*Examples*
S	Speaker	"the one nearest myself" "the cage closest to me"
A	Addressee	"info booth on your left" "2nd one directly in front of you"
L	Landmark	"the one farthest from the grape press" "the sandbox closer to the jungle gym"
C	Compass	"the west one" "the eastern-most barn"
S&L	Speaker & Landmark	"start at the iguana exhibit and walk towards me. The rat is in the snake cage on your right" "back to cabinet, facing me, on left"
S&C	Speaker & Compass	"couch southwest of me" "entrance due east of me"
A&L	Addressee & Landmark	"right of the grape press" [from A's perspective] "go to the display then keep right"
A&C	Addressee & Compass	"the entrance southwest of you" "go southeast"
L&C	Landmark & Compass	"if the ticket booth is north of you, it's in the tent to your left" "southwest of booth"

Table 3

Overall Use of Perspective Elements in Participants' Initial Descriptions, Experiment 1

Perspective element	Mean frequency per participant		Maximum possible frequency	% of maximum frequency	
	US	Japan		US	Japan
Speaker	1.8	1.9	20	9	10
Addressee	15.2	13.9	20	76	70
Landmark	3.3	4.5	12	28	38
Compass	6.0	5.1	12	50	42

Table 4

Perspective Use by Scenario Type, Experiment 1

A. Copresent Scenarios

Scenario type	Culture	Mean % respondents using perspective					
		S	S&N	A	A&N	N	Other
Baseline	US:	8		88			4
	Japan:	12		88			1
Landmark/equidistant	US:	9	1	73	14	1	1
	Japan:	11	0	76	5	8	0
Landmark/skewed	US:	7	0	61	8	23	1
	Japan:	8	0	51	1	40	0
Compass/north-up	US:	9	0	58	5	27	0
	Japan:	11	0	57	5	26	1
Compass/north-side	US:	3	1	60	6	28	1
	Japan:	6	0	59	5	30	0

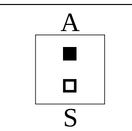
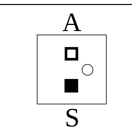
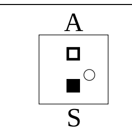
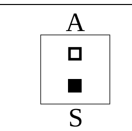
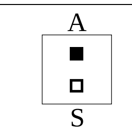
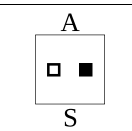
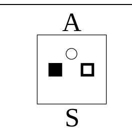
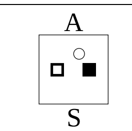
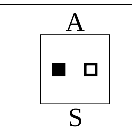
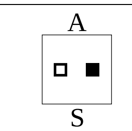
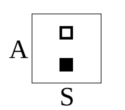
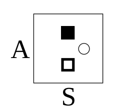
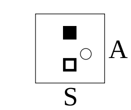
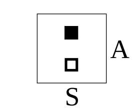
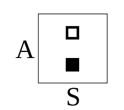
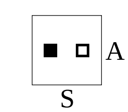
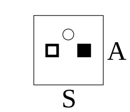
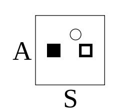
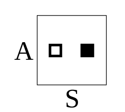
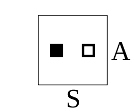
B. Remote Scenarios

Culture	Mean % respondents using perspective			
	C	L	C&L	Other
US	65	17	17	2
Japan	43	39	18	0

Note. See Table 2 for perspective codes. N stands for perspectives of type L or C, depending on the scenario type.

Table 5

Percent of Participants Initially Using Speaker Elements in Copresent Scenarios, Experiment 1

Config. code	Scenario type: Baseline		Landmark/ equidistant		Landmark/ skewed		Compass/ north-up		Compass/ north-side	
	US	Japan	US	Japan	US	Japan	US	Japan	US	Japan
FN										
	0	3	33	33	23	19	29	30	0	1
FL										
	6	3	4	1	1	4	3	1	0	3
PN										
	21	27	3	4	4	6	3	7	13	11
PL										
	4	16	3	4	1	1	3	10	3	9

Note. A and S represent icons for the addressee (Agent Z) and the participant (Agent U), respectively; the squares represent containers, the target filled and the non-target open; and the circles represent landmarks.

Table 6

Overall Use of Perspective Elements in Participants' Initial Descriptions, Experiment 2

Perspective element	Mean frequency per participant		Maximum possible frequency	% of maximum frequency	
	US	Japan		US	Japan
Speaker	10.0	10.0	20	50	50
Addressee	3.7	3.8	20	18	19
Landmark	4.9	5.6	12	24	28
Compass	8.2	7.0	12	41	35

Table 7

Perspective Use by Scenario Type, Experiment 2

A. Copresent Scenarios

		Mean % respondents using perspective					
Scenario type	Culture	S	S&N	A	A&N	N	Other
Baseline	US:	66		31			3
	Japan:	65		31			4
Landmark/equidistant	US:	56	6	23	0	10	4
	Japan:	57	2	15	3	21	1
Landmark/skewed	US:	37	4	14	0	44	1
	Japan:	36	1	9	1	53	0
Compass/north-up	US:	34	12	1	3	50	1
	Japan:	38	10	8	4	39	1
Compass/north-side	US:	24	5	9	3	57	1
	Japan:	28	8	10	7	45	1

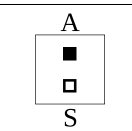
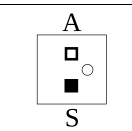
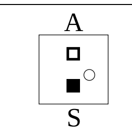
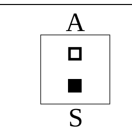
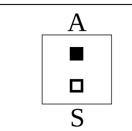
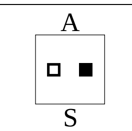
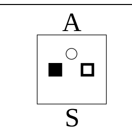
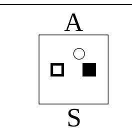
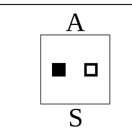
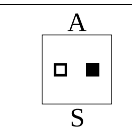
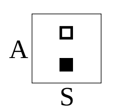
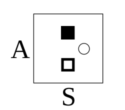
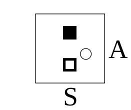
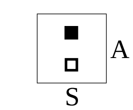
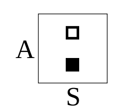
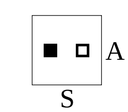
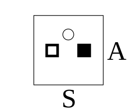
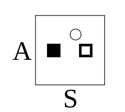
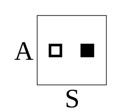
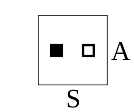
B. Remote Scenarios

	Mean % respondents using perspective			
Culture	C	L	C&L	Other
US	45	28	26	0
Japan	41	39	19	0

Note. See Table 2 for perspective codes. N stands for perspectives of type L or C, depending on the scenario type.

Table 8

Percent of Participants Initially Using Agent Z Elements in Copresent Scenarios, Experiment 2

Config. code	Scenario type: Baseline		Landmark/ equidistant		Landmark/ skewed		Compass/ north-up		Compass/ north-side	
	US	Japan	US	Japan	US	Japan	US	Japan	US	Japan
FN										
	88	74	4	4	0	0	0	2	29	44
FL										
	4	9	4	11	0	0	4	9	4	4
PN										
	12	19	29	37	12	20	8	20	4	13
PL										
	29	22	67	22	42	20	8	17	17	7

Note. A and S represent icons for Agent Z and the participant (Agent U), respectively; the squares represent containers, the target filled and the non-target open; and the circles represent landmarks.

Table 9

Overall Use of Perspective Elements in Participants' Initial Descriptions, Experiment 3

Perspective element	Mean frequency per participant		Maximum possible frequency	% of maximum frequency	
	US	Japan		US	Japan
Speaker	3.4	3.0	24	14	13
Addressee	8.9	6.0	24	37	25
Landmark	9.2	10.6	20	46	53
Compass	7.8	9.4	16	49	59

Table 10

Perspective Use by Scenario Type, Experiment 3

A. Copresent Scenarios

		Mean % respondents using perspective					
Scenario type	Culture	S	S&N	A	A&N	N	Other
Baseline	US:	31		69			0
	Japan:	31		68			0
Landmark	US:	13	0	38	0	50	0
	Japan:	11	0	21	0	67	0
Compass	US:	15	0	34	4	47	0
	Japan:	12	0	24	0	64	0
Landmark & Compass	US:	7	0	22	2	69	0
	Japan:	5	0	6	0	88	0

B. Remote Scenarios

	Mean % respondents using perspective			
Culture	C	L	C&L	Other
US	42	40	18	0
Japan	47	39	14	0

Note. See Table 2 for perspective codes. N stands for perspectives of type L or C (or L&C), depending on the scenario type.

Table 11

Spatial Relation Choice within "Pure" Perspectives, Experiment 3

A. Speaker, Addressee, and Landmark Perspectives

		% Responses Using Relation			
Perspective	Culture	Near	Far	Left	Right
Speaker	US:	71.0	2.3	16.8	9.8
	Japan:	76.7	0.6	5.7	17.0
Addressee	US:	59.4	0.9	20.5	19.2
	Japan:	65.6	0.6	6.4	27.4
Landmark	US:	98.4	0.6	0.6	0.4
	Japan:	99.6	0.2	0.2	0.0

B. Compass Perspectives

		% Responses Using Relation			
Perspective	Culture	North	South	East	West
Compass	US:	50.1	8.7	22.6	18.5
	Japan:	55.8	1.5	27.2	15.5

Table 12

Spatial Relation Pairings in Experiment 4

Speaker's perspective		Addressee's perspective	
Relation	Phrase	Relation	Phrase
near	"nearest to me"	*far*	"farthest from you"
near	"nearest to me"	*right*	"to your right"
far	"farthest from me"	*near*	"nearest to you"
far	"farthest from me"	*left*	"to your left"
left	"to my left"	*near*	"nearest to you"
left	"to my left"	*right*	"to your right"
right	"to my right"	*far*	"farthest from you"
right	"to my right"	*left*	"to your left"

Table 13

Proportion of Participants Choosing Addressee's Perspective

Location of target, from speaker's perspective	Location of target, from addressee's perspective			
	near	*left*	*right*	*far*
near	–	–	**26%**	**10%**
left	**96%**	–	50%	–
right	–	44%	–	**70%**
far	**93%**	36%	–	–

Note. Boldface values differ from 50% with $p < .05$ (sign test). "–" indicates cases that were not tested.

Figure Captions

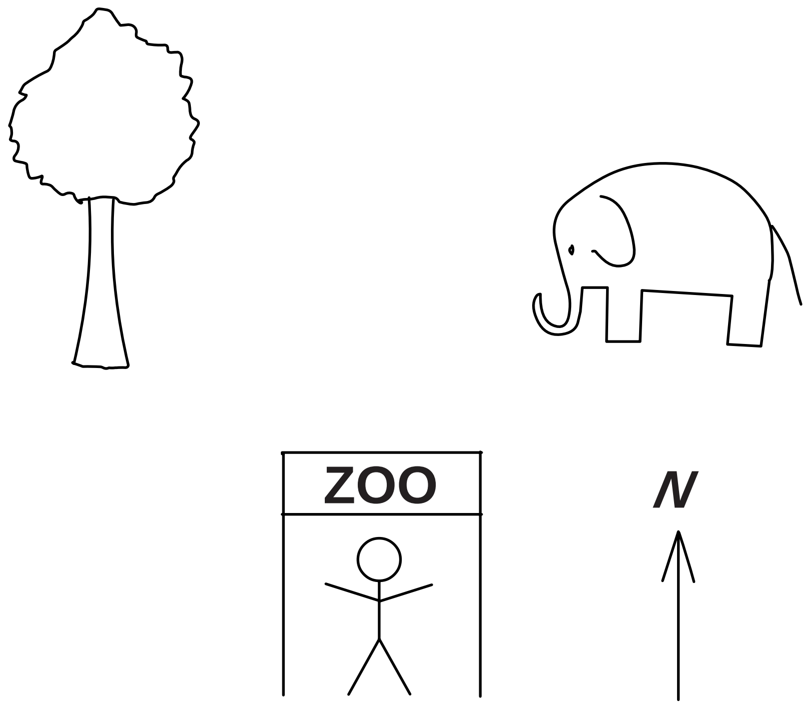
Figure 1. A simple spatial scene.

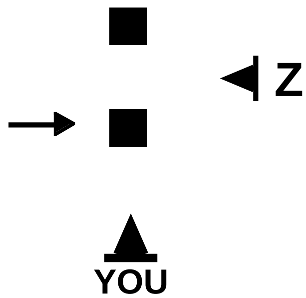
Figure 2. Diagram of a baseline situation. The location of the target (indicated by arrow) is presumed to be described more easily from "your" perspective than from Z's.

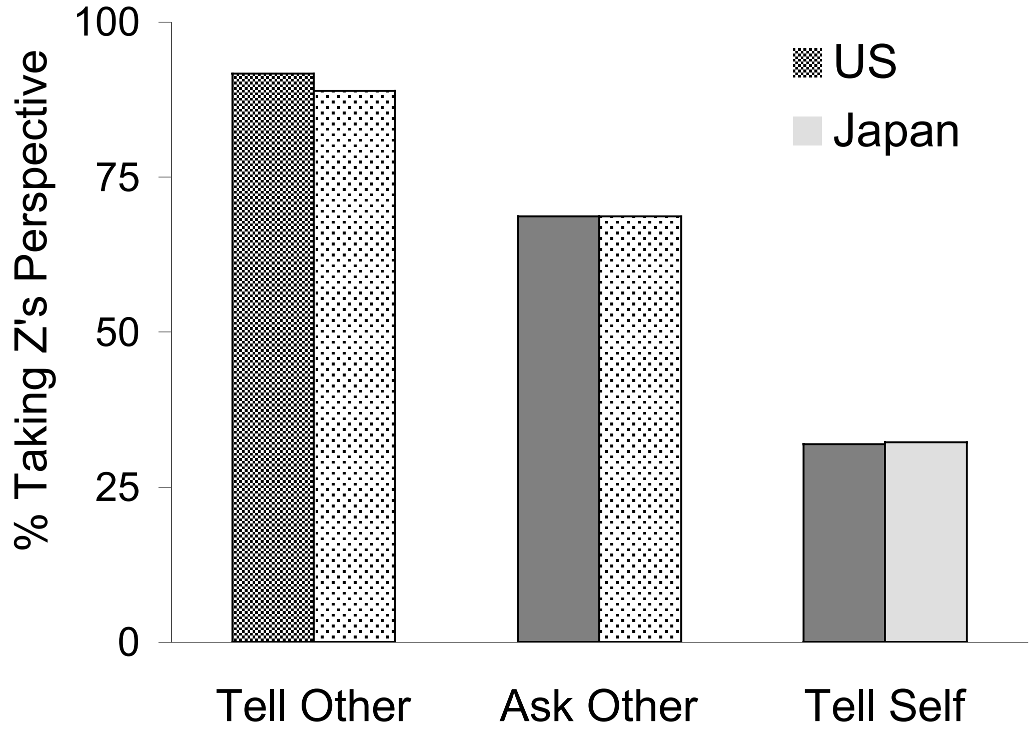
Figure 3. For each of the three experiments, probability that participant adopted Agent Z's perspective in a baseline scenario (in which only alternative was to use the speaker's perspective). Experiments are ordered in decreasing task demand on Agent Z.

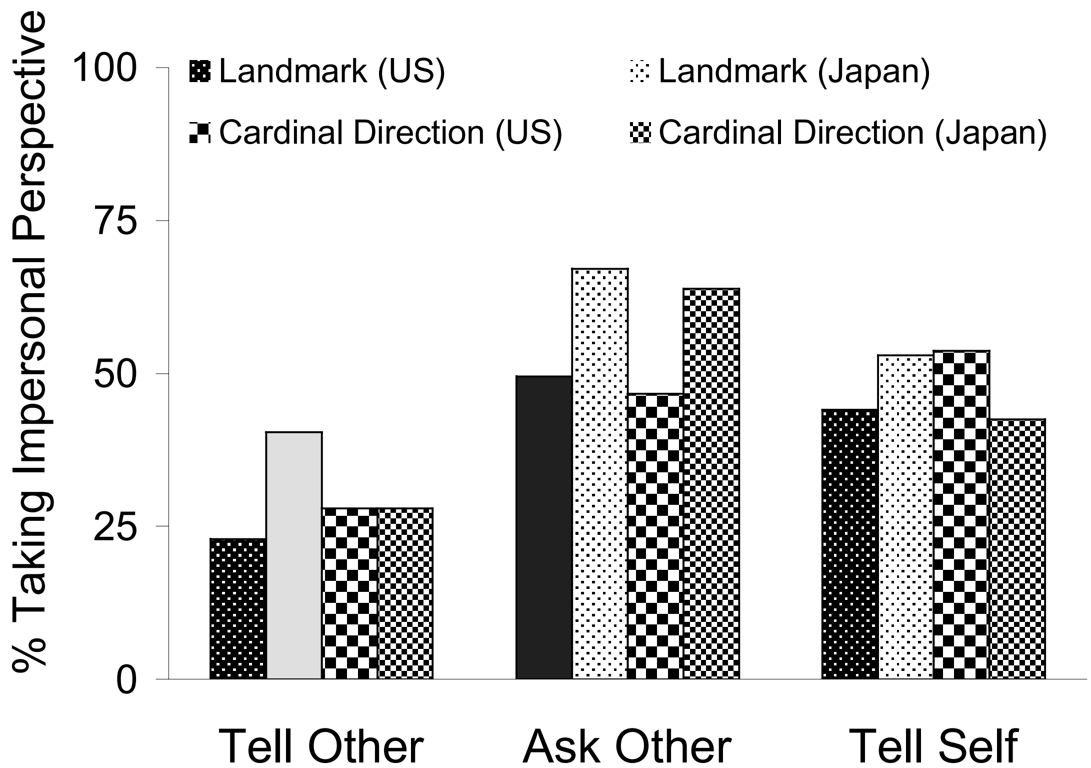
Figure 4. For each of the three experiments, probability that participant adopted a neutral perspective in a scenario in which the choice was to take a personal perspective (the speaker's or Agent Z's) or a neutral perspective (based on landmarks or cardinal directions). Experiments are ordered in decreasing task demand on Agent Z.

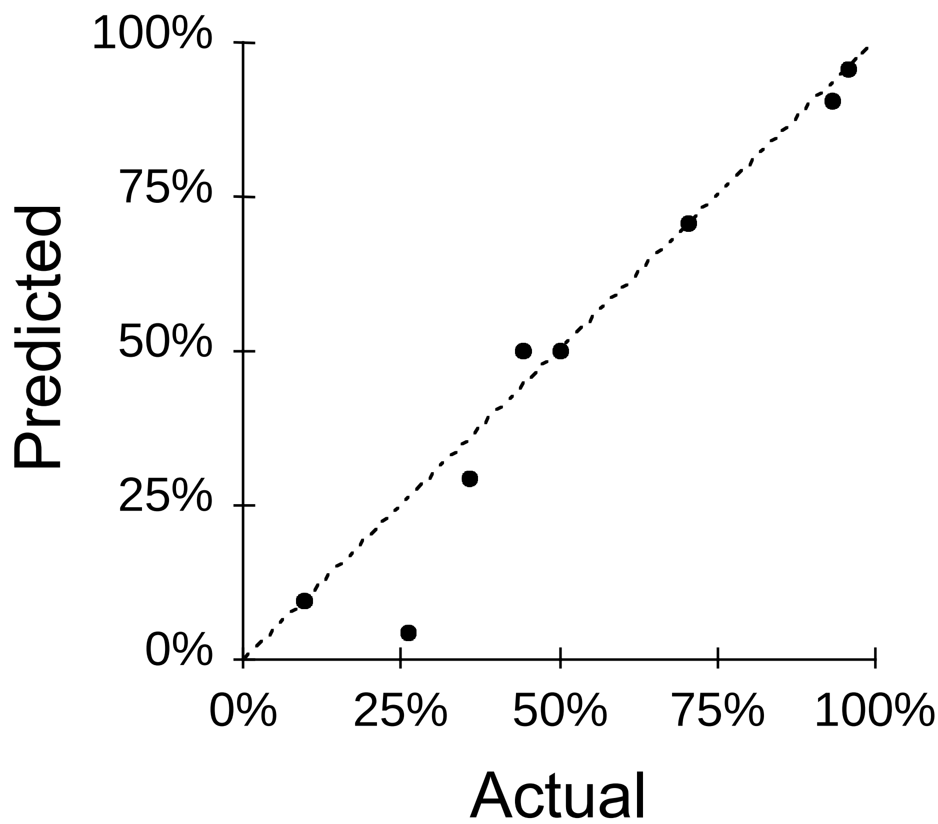
Figure 5. Actual vs. predicted proportion of participants choosing description from addressee's perspective.


ZOO
N


Z
YOU


100
75
50
25
0
% Taking Z's Perspective
US
Japan
Tell Other
Ask Other
Tell Self


% Taking Impersonal Perspective
100
75
50
25
0
Landmark (US)
Landmark (Japan)
Cardinal Direction (US)
Cardinal Direction (Japan)
Tell Other
Ask Other
Tell Self


100%
75%
50%
25%
0%
Predicted
0%
25%
50%
75%
100%
Actual