Spatial Mental Models from Descriptions

Barbara Tversky
Department of Psychology, Building 420, Stanford University, Stanford, CA 94305-2130

Nancy Franklin
State University of New York, Stony Brook, NY 11794-2500

Holly A. Taylor
Tufts University, Medford, MA 02155

David J. Bryant
Northeastern University, Boston, MA 02115

Spatial language is widely used, both literally to describe space and figuratively to express a broad range of ideas. This article reviews two projects studying the nature of mental representations of space induced entirely by language. The first project investigates perspective in descriptions of large-scale (e.g., convention center, town) space. People typically describe environments using a route or survey perspective, or a mixture of both. Route perspectives take a view from within the environment and describe the locations of landmarks with respect to a moving observer in terms of the observer's left, right, front, and back. Survey perspectives take a view from above the environment and describe locations of landmarks with respect to each other in terms of north, south, east, and west. Features of the environment, such as having a single or multiple path, affect choice of perspective. In comprehension, readers seem to form the same perspective-free mental representation irrespective of description perspective. They respond with equal speed and accuracy to inference questions from either perspective regardless of read perspective. The second project investigates mental representations of the objects located immediately around the body. Readers seem to form mental spatial frameworks, extensions of the three body axes, associating objects to the frameworks. The accessibility of the three axes depends on characteristics of the body, characteristics of the perceptual world, and posture of the observer. For example, for an upright observer, times to access objects along the head/feet axis are fastest because it is an asymmetric axis of the body and is correlated with the only asymmetric axis of the world, that created by gravity. Times to access objects along the front/back axis are next fastest as it is an asymmetric axis of the body, and times to the left/right axis are slowest as it is an axis with few asymmetries. Evidence for the spatial framework hypothesis was obtained in a variety of situations, varying posture, perspective, number of observers, and cause of reorientations.

Introduction

There are many situations in which we have no choice but to explain things in words rather than show them. Although not perfect, language often serves such purposes quite well. One of those situations is describing environments, whether we are telling a new friend about the neighborhood where we grew up, or an old friend about our recent trip abroad, or a stranger how to find their way to the campus bookstore. Novels, tourist guides, and history and science texts abound in spatial descriptions. A number of years ago, we began exploring the nature of spatial mental representations produced by words alone.

Languages are rich in spatial vocabulary probably because space is so important to every aspect of our lives. Spatial language has been co-opted for other uses and pervades speech. We change *perspectives* and enter new *fields,* careful not to encroach on another's *space.* Some spatial information seems easily conveyed by language, and other information less so. Those terms that have proved their usefulness by being old and frequent and metaphorically extended are the ones that should communicate effectively. A notable example is the terms that describe spatial relations, such as *front, back, left, right, above, below,* and *north, south, east,* and *west.* Part of the communicative effectiveness of these terms is shared knowledge about the appearances of environments. Other spatial information seems more difficult to con-

This research was supported by the Air Force Office of Scientific Research, Air Force Systems Command, USAF, under Grant or Cooperative Agreement Number AFOSR 89-0076.

 CCC 0002-8231/94/090656-13

vey, for example, precise information about distance and angle. Conveying distance and angle requires recently developed technical language that most of us do not use frequently enough to become accurate. Distance information that relies on shared knowledge about stereotyped units, such as rooms or blocks in American culture, should be better conveyed and understood than metric information.

Our own work has investigated communication of spatial relations for large and local environments. The work of Johnson-Laird and his collaborators (Ehrlich & Johnson-Laird, 1982; Mani & Johnson-Laird, 1982); of Glenberg, Meyer, and Lindem (1987); and of Perrig and Kintsch (1985) encouraged us in this endeavor. We were also encouraged by the related work of Morrow, Greenspan, and Bower (1987) and Morrow, Bower, and Greenspan (1989) showing that stereotypic distance units may be reflected in sentence comprehension time. Globally, we have been interested in the nature of the mental representations that readers construct from descriptions, and have used reaction time and accuracy measures as reflections of the mental representations. The project has taken us into a number of other issues as well, including perspective and perspective-switching, the nature of spatial descriptions, and the comparison of environments learned by description to those learned by perception. We will begin with the work on large environments.

Comprehending Route and Survey Descriptions of Large Environments

Introduction

An informal survey of tourist guides revealed that tourist sites seem to be described from one of two perspectives, survey or route. In a *survey* perspective, descriptions take a view from above the environment and describe the locations of landmarks relative to one another in terms of north, south, east, and west. In a *route* perspective, descriptions take a view from within the environment, and take addressees on a mental tour of the environment describing the locations of landmarks relative to the changing position of the addressee in terms of the addressee's front, back, left, and right. These two perspectives correspond to the two major ways of learning about environments, from maps and from navigation. The route perspective corresponds to what linguists have called a *deictic* perspective and the survey perspective corresponds to what linguists have called an *extrinsic* perspective (Fillmore, 1975; Miller & Johnson-Laird, 1976).

Taylor and Tversky (1992b) wondered whether the two perspectives led to different mental representations, that is, whether perspective is encoded in the spatial mental representations of the described environments. Previous work on descriptions and on learning environments suggested that perspective would be encoded in the mental representations of environments. Readers remember information associated with the narrative perspective better than information associated with an alternative perspective (Abelson, 1975; Anderson & Pichert, 1978; Bower, 1978). Some aspects of environments, such as relative directions, are better acquired or more accessible from studying maps, and other aspects of environments, such as travel distance, are better acquired from navigation (Evans & Pezdek, 1980; Sholl, 1987; Streeter, Vitello, & Wonsiewicz, 1985; Thorndyke, 1981; Thorndyke & Hayes-Roth, 1982). The research on which our own was fashioned, experiments by Perrig and Kintsch (1985), also found differences due to perspective.

Design

We first designed four environments, varying in area, each with about a dozen landmarks. The largest environment was a county-sized recreation area and the next largest was a small town. The two smaller environments were a zoo and a convention center. For each environment, we wrote two descriptions, one from a route perspective and one from a survey perspective. Each description contained information locating the landmarks in the environment according to the appropriate perspective. To make the descriptions more interesting and realistic, each also contained identical nonlocative information, for example, describing activities that took place at the landmarks. We pretested the descriptions to make sure that they were equally coherent and that each allowed readers to accurately place all the landmarks. The survey and route descriptions of the convention center are displayed in Tables 1 and 2 as examples. In four experiments, subjects studied a route or a survey description of an environment, followed by a number of tests of memory. First, subjects verified true–false statements, both locative and nonlocative, from the descriptions. The locative statements were either verbatim from one of the texts, route or survey, or were inference statements from one of the two perspectives. The inference statements contained information that could be inferred from information in the descriptions, but had not been explicitly stated in either description. Following the verification of statements, subjects drew a map of the environment.

Results

From reading either narrative perspective, subjects drew highly accurate maps of the environments, indicating that they formed excellent spatial mental representations from the descriptions. If perspective is encoded in the spatial mental representations formed from the descriptions, then responses to inference statements from the same perspective as the narrative should be faster and

TABLE 1. Convention center: Survey perspective.

Several companies that manufacture electronics have decided to get together for a convention to show their wares. A large convention center was chosen because its large rectangular floor plan can be easily changed to accommodate the needs of various conventions. Temporary wall dividers are used to separate the displays and to form a single entrance to each display. The displays have been grouped according to three categories—Visual Equipment, Personal Computers, and Audio Equipment. The rectangular center section of the building is divided into four displays for the visual equipment. In the northwest corner of the center section, with the entrance facing north, are the Televisions. Like many television displays, the sets are lined up along the walls, all tuned to the same station. In the northeast corner of the center section, with the entrance facing north, are the VCRs. In the southwest corner of the center section, with the entrance facing south, are the 35mm Cameras. In the southeast corner of the center section, with the entrance facing south, are the Movie Cameras. The Movie Cameras are set up to film people as they walk by the display. The remainder of the displays are along the outer, rectangular wall of the Convention Center. The east wall has only one display, the Personal Computers. This display is in the northeast corner and extends for about half of the east wall. There are software samples available for potential customers to test the various computers. Along the north wall are the two Audio Equipment displays—the Stereo Components and the CD Players. Along the north wall, directly west of the Personal Computers, are the Stereo Components. This display includes such items as receivers, turntables, speakers, and tape decks. Directly west of the Stereo Components are the CD Players. In addition to the displays, there are four permanent features of the Convention Center located along the west and south walls—the Cafeteria, the Restrooms, the Office, and the Bulletin Board. Just west of the CD Players, beginning in the northwest corner of the Convention Center and extending for about half of the west wall, is the Cafeteria. The Cafeteria is privately run by a family that leases the space on a permanent basis from the Convention Center. Directly south of the Cafeteria, on the west wall, are the Restrooms. Directly south of the Restrooms, extending from the southwest corner for about a third of the south wall, is the Office. East of the Office, covering about half of the south wall, is the Bulletin Board. The Bulletin Board is used in every convention for the business cards of the participating companies. East of the Bulletin Board, on the east side of the building near the southeast corner, is the entrance.

more accurate than to inference statements from the other perspective. However, this failed to happen in four experiments, including an experiment where subjects read only a single description and did not know they would be asked to draw maps. This is evidence that perspective is not encoded in the spatial mental models formed from these types of descriptions. Subjects were faster and more accurate to verbatim statements than to inference statements. Altogether, survey and route perspectives were equally effective in conveying information about the environments, and route and survey questions were comparable in difficulty.

Interpretation

Readers seemed to have formed at least two mental representations of the descriptions. First, they formed a representation of the language of the description that allowed them to respond faster and more accurately to the verbatim statements than to the inference statements. Second, they formed a representation of the situation described by the text, what we have called (after Johnson-Laird, 1983) a spatial mental model, that allowed them to draw accurate maps of the environments and to respond to the inference statements.

The spatial mental models subjects formed of the described environments were apparently indifferent to perspective. They allowed subjects to respond as quickly and accurately to statements in the same perspective that they had read and in the other perspective, for both perspectives. This suggests that the spatial mental models contained information about the spatial relations among landmarks in a perspective-free fashion, allowing the taking of either perspective with equal ease. As such, spatial mental models are similar to structural descriptions, common in computer models of object recognition, and to architect's models of buildings. Each incorporates the spatial relations among parts and allows the taking of many perspectives.

Caveats and Qualifications

We do not mean to claim that all spatial mental representations of environments, whether learned from de-

TABLE 2. Convention center: Route description.

Several companies that manufacture electronics have decided to get together for a convention to show their wares. A large convention center was chosen because its large, rectangular floor plan can be easily changed to accommodate the needs of various conventions. Temporary wall dividers are used to separate the displays and to form a single entrance to each display. The displays have been grouped according to three categories—Visual Equipment, Personal Computers, and Audio Equipment. You go to the east side of the building near the southeast corner where you find the entrance. As you walk into the building, you see, on your left, a Bulletin Board. The Bulletin Board is used in every convention for the business cards of the participating companies. Continuing straight ahead from the entrance, where the Bulletin Board is on your left, you reach, on your right, the Movie Cameras. The Movie Cameras are set up to film people as they walk by the display. Walking past the Movie Cameras on your right, you see, again on your right, the 35mm Cameras. On your left, stretching into the corner of the building, is the Office. From the Office, you are forced to turn right and you see, to your immediate left, the Restrooms. You continue forward from the Restrooms until you see, on your left stretching into the corner of the building, the Cafeteria. The Cafeteria is privately run by a family that leases the space on a permanent basis from the Convention Center. From the Cafeteria, you walk forward, until you are forced to turn right and you see, to your immediate left, the CD Players. On your right are the Televisions. Like many television displays, the sets are lined up along the walls, all tuned to the same station. You walk past the Televisions, on your right, and continue forward until you see, again on your right, the VCRs. On your left are the Stereo Components. This display includes such items as receivers, turntables, speakers, and tape decks. From the Stereo Components you walk forward until you are forced to turn right and you see, to your immediate left, the Personal Computers. There are software samples available for potential customers to test the various computers. From the Personal Computers, you walk until you reach, on your left, the corridor leading to the entrance of the building.

scriptions or learned from experience, have this character. These experiments were done under ideal conditions. The environments and the descriptions were carefully composed, the memory load was light, the time given for learning was ample. If these conditions are not met, then there is no reason to expect that subjects will form coherent and complete mental representations of the spatial relations that are perspective-free. In fact, many experiments have demonstrated just that, that people's mental representations of environments are often incoherent and incomplete and from a particular perspective. It is important to know that this need not be the case. But it is also important to remember that excellent as they were, the spatial mental models subjects formed did not contain metric information, only categorical spatial relations. The descriptions did not relay metric information, and we do not think that precise metric information is easy to relay accurately.

Producing Descriptions of Large Environments

Introduction

After completing this research, we wondered what perspectives people actually took in describing environments. Our own descriptions had been carefully and self-consciously fashioned, so we wanted to know how laypeople spontaneously constructed descriptions. There was a widespread belief in the psycholinguistics literature that people ordinarily give route descriptions of environments (cf. Levelt, 1982). This belief came primarily from a well-known study of descriptions of New York City apartments by Linde and Labov (1975). In that study, the overwhelming majority of descriptions took listeners on a mental tour of the environment. More recent work by Ehrich and Koster (1983), Levelt (1982), and Ullmer-Ehrich (1982) strengthened that conclusion. The conclusion was justified by an analysis of Levelt's (1989), in which he pointed out that space is multidimensional but speech is linear. Because a route is similar to exploration of an environment, it provides a natural way to linearize space. Survey perspectives are also natural ways of conceptualizing environments. People get survey perspectives from heights, and throughout time, cultures have repeatedly invented and used maps, which give survey perspectives.

Investigating a Greater Variety of Environments

Altogether, the range of environments studied thus far seemed too narrow to come to any definitive conclusions about perspective. We embarked on a mirror-image of our first set of tasks (Taylor & Tversky, 1992a, 1993). We constructed maps of the recreation area and convention center from our previous work, and added to that a map of an amusement park. Subjects studied one of these maps. Later, they wrote descriptions of the environments from memory. Their descriptions were coherent and accurate, allowing a naive group of subjects to place nearly all the landmarks correctly.

Subjects produced survey, route, and a mixture of survey and route descriptions, with the majority of the descriptions mixed, followed by survey, followed by route. The town received mainly survey and mixed perspective descriptions, with very few route descriptions, and the convention center received mainly route and mixed perspective descriptions, with very few survey descriptions. This suggested that characteristics of the environment determined the perspective chosen. These findings corroborate the results of the previous set of studies. They suggest that perspective is not inherent in people's mental representations of environments. Selection of a description perspective, then, has to do with how convenient it is to describe that environment with that perspective. Convenience of perspective seems to depend at least in part on characteristics of the environments themselves.

The town and the convention center were described with different perspectives and differed on a number of environmental characteristics that might be relevant to perspective. The town was an open environment, whereas the convention center was enclosed; the town was relatively large, and the convention center was small; the town had landmarks on several size scales, mountains, river, roads, and buildings, and the convention center had landmarks on a single size scale, rooms; the town had more than one possible route and the convention center had only a single route.

Varying Environmental Characteristics

In the next experiment (Taylor & Tversky, 1993), we tested the hypothesis that characteristics of the environment affect perspective choice by systematically varying the environments on a number of features that differentiated the town and the convention center, and that might be related to choice of perspective. We designed 16 different environments, crossing the four characteristics of the environments. Subjects wrote descriptions of four of the environments from memory. Forty-five percent of the descriptions had a survey perspective, 34% had a mixed perspective, and 21% had a route perspective. Two of the environmental variables had no effect on choice of perspective: whether the environment was open or enclosed, and whether the environment was of a relatively large area, such as a county, or of a relatively small area, such as a museum.

Two other variables did affect the selection of a perspective. There were fewer mixed and more route descriptions when the environments had only a single path and when the landmarks were on the same size scale. This makes sense in terms of the pragmatics of constructing a coherent description. When an environment has only one path through it rather than many possible paths, it is easier to construct a route. The presence of landmarks on the same size scale makes it more likely that each landmark will be described in terms of the adjacent landmark. When one or more landmarks are larger and more salient, then it is more likely that smaller landmarks will be described with respect to the large landmark. If many landmarks are described with respect to a single salient one, it seems easier to differentiate them by their canonical directions from the large landmark than by taking a tour around the large landmark.

Varying Acquisition Characteristics

Another difference between our studies and the previous work is that in our studies, subjects learned the environments by studying maps, whereas in some of the previous work, subjects had learned the environments by navigating them. Linde and Labov's subjects described apartments they had lived in and presumably walked around in quite a bit. Studying a map might encourage a survey perspective whereas navigation might encourage a route perspective. Acquisition conditions can only explain part of the results, however. Levelt's (1982) subjects described a map of a network of colored dots, and produced route descriptions.

We ran a small study where we asked students to describe environments that they presumably experienced primarily by navigation rather than maps, two areas on campus and their neighborhoods at home. The descriptions were not as detailed or coherent or accurate as the descriptions of the memorized maps. Most of them would not have allowed a naive subject to correctly place most of the landmarks. In this study, the majority of the descriptions were route descriptions, in contrast to our previous studies. However, subjects did produce survey and mixed perspective descriptions of environments they had experienced primarily or only by navigation.

Determinants of Description Perspective

Selection of perspective to describe an environment is clearly affected by characteristics of the environment, but it may also be affected by acquisition conditions.

Organization of Descriptions

Taylor and Tversky (1992a) also examined the organization of subjects' descriptions of environments. Globally, the descriptions were organized hierarchically. In some instances the hierarchy was perceptually determined, that is, larger and more salient features were described first. For other environments, the hierarchy was determined functionally, for example, beginning with an entrance. On the local level, the descriptions were organized by referencing devices. Except for introductory sentences, most of the description statements referred back to something mentioned in a previous statement. The introductory statements and many of the intermediary statements anticipated information in subsequent statements. Overlapping references linked the sentences locally, as in a chain.

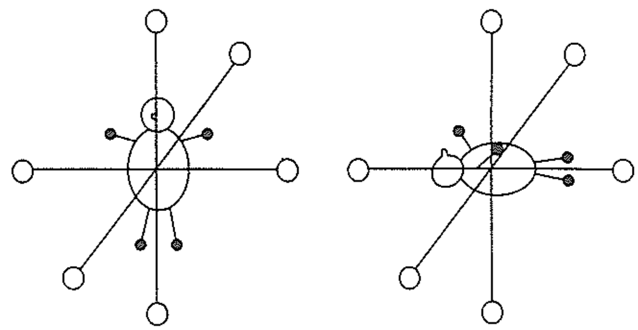

FIG. 1. Internal spatial framework. (a) Upright observer. (b) Reclining observer.

Local Environments: Spatial Frameworks

Introduction

The experiments on descriptions of large environments showed that under ideal circumstances people can form spatial mental representations of environments that are coherent, complete, and perspective-free. Responding to particular questions, however, required taking a specific perspective on the environment. We turn now to discuss experiments investigating responding to a specific perspective, namely that of an observer in a scene. At the same time, we switch from large to local environments. We also switch referents of the term "we." Here, "we" refers initially to Franklin and Tversky, and later includes Bryant as well.

Task

The situation we (Franklin & Tversky, 1990) chose to study is one that people find themselves in most of the time, in a setting, surrounded by objects. We were interested in the simplest variant of that scene, keeping track of the positions of objects under the simplest form of navigation, turning in place. It is a task that people seem to do effortlessly. In order to study it, we wrote narratives describing "you," the observer and reader, in a setting such as a barn, with objects such as a lantern, a pail, a rake, and a saddle located beyond "your" head, feet, front, back, left, and right. A schematic representation of this situation, that of an upright observer, appears in Figure 1. Subjects studied the narratives until they knew them well, and then turned to a computer that oriented them toward one of the objects, and queried them for the objects located in each of the six directions beyond the body. When all locations had been probed, the computer again reoriented the reader, and again queried the reader for the objects now located in the six directions beyond the body. Performance was essentially error-free, so the data of interest are the reaction times to each of the six directions.

Equiavailability and Mental Transformation Accounts Do Not Work

How might someone in this situation perform the task? We considered three possible theories. According to the first theory, the *equiavailability* theory, all directions in space are in principle equally salient and available, much like viewing a picture, where certain objects in the picture may attract attention more than others, but not because of the direction per se. Equiavailability predicts equal reaction times to all directions. The second theory extends the research in mental imagery (for reviews, see Finke & Shepard, 1986; Kosslyn, 1980) from the typical two-dimensional setting to the current three-dimensional setting. According to a *mental transformation* account, readers would imagine the scene and themselves in it. To verify the object in a given direction, readers would imagine themselves turning to inspect that direction. This theory predicts that times should be fastest to front, next fastest to directions 90 degrees from front, that is, left, right, head, and feet, and slowest to back, which is displaced 180 degrees and requires the longest mental transformation. The data of the first five experiments, and by now, many more, reject both the equiavailability and mental transformation theories as accounts for the standard situation (something like the equiavailability account seems to hold for certain complex situations; see Franklin, Tversky, & Coon, 1992).

Spatial Framework Theory: Upright Case

The theory that accounts for the data is the *spatial framework theory.* This theory is based on analyses of language and space by Clark (1973), Fillmore (1975), Levelt (1984), Miller and Johnson-Laird (1976), and Shepard and Hurwitz (1984), but it is different from any of the previous analyses. According to it, readers construct a mental spatial framework from extensions of the body axes, and associate objects to the appropriate axes. Readers update the observer's positions as the observer is reoriented. The three axes vary in accessibility depending on characteristics of the body, characteristics of the perceptual world, and the posture of the body. For the upright observer, the most salient and accessible axis is the head/feet axis. This is because it is an asymmetric axis of the body and it correlates with the only constant asymmetric axis of the perceptual world, the up–down axis induced by gravity. The next most accessible axis is the front/back axis, which has perceptual and functional asymmetries, and the least accessible axis is the left/right axis, which has no salient asymmetries. The pattern of reaction times in the four upright experiments of Franklin and Tversky (1990) and of subsequent experiments conformed to this pattern, fastest reaction times to head/feet, then front/back, then left/right.

Spatial Framework Theory: Reclining Case

When the observer reclines and reorients by turning onto the front, back, left, and right sides, the situation changes. See Figure 1 for a schematic representation of this situation. Gravity no longer corresponds to any major axis of the body, so the accessibility of axes relies solely on characteristics of the body. Both front/back and head/feet axes have biological asymmetries, but the front/back asymmetry is more influential. The front/back axis separates the world that can be perceived and manipulated from the world that cannot be easily perceived or easily manipulated. Thus, according to the spatial framework theory, for a reclining observer, fastest times should be to the front/back axis, followed by the head/feet axis, and, last, the left/right axis. This pattern of reaction times emerged for the two reclining experiments of Franklin and Tversky (1990), and for subsequent research as well. Both the upright and reclining patterns were replicated in an experiment using objects as probes for directions (Bryant & Tversky, 1992). In all of the experiments comparing upright and reclining postures, reaction times in the reclining condition were longer than those in the upright condition. This finding fits with the premises of the spatial framework theory. In their interactions in the world, people typically are upright as they turn and navigate.

Spatial Framework Theory: Third Person Narratives

After examining the simplest case, of an observer surrounded by objects, standing or reclining in the environment, and turning to face different objects, we began to vary the situation. In the first variation, we described the scenes in the third person rather than the second person. Thus, we have substituted an egocentric or deictic frame of reference with an object-centered or intrinsic frame of reference (e.g., Levelt, 1984; Marr & Nishihara, 1978). We expected that readers would still adopt the internal perspective of the observer in the scene, even when that observer was described as a person other than "you" (Bryant, Tversky, & Franklin, 1992). Thus, subjects would turn an intrinsic situation into a deictic one, perhaps because that is the situation most familiar to us, hence easiest for us. The same pattern of reaction times resulted, suggesting that subjects did in fact adopt the perspective of the third person.

We then replaced the human observer with an inanimate object that was successively turned to face different objects in the environment (Bryant et al., 1992). We chose objects that had intrinsic fronts, backs, tops, bottoms, lefts, and rights, such as a saddle. Objects do not