Mental Models in Discourse Processing and Reasoning
G. Rickheit and C. Habel (Editors)
© 1999 Elsevier Science B.V. All rights reserved.

11

Mental Models of Spatial Relations and Transformations from Language

Barbara Tversky, Joseph Kim, Stanford University, Stanford, CA, USA
Andrew Cohen, Indiana University, IN, USA

When we read a vivid description of a social encounter or a journey, we often are able to imagine where the characters and objects are in the scene and how they move about as the action proceeds. Researchers studying how people use and understand language have come to agree that language comprehension entails construction of a situation model or mental representation of the situation described by the discourse in addition to mental representations of the words and sentences (e.g., Bransford *et. al.*, 1972; van Dijk and Kintsch, 1983; Mani and Johnson-Laird, 1982). At about the same time, researchers studying spatial cognition began to investigate the power of language alone to instill complex internal representations of space (e. g., Ehrlich and Johnson-Laird, 1982; Tversky, 1991b). The confluence of these fields has enriched both. For example, language researchers have often assumed that people's mental representations of space are Euclidean-like and veridical, for example, that distance information is represented analogically (e.g., Wilson *et al.*, 1993). However, a long history of research in spatial cognition has demonstrated that people's mental representations of space are qualitative and piecemeal, thus likely to be inconsistent and distorted (e.g., Hirtle and Jonides, 1985; Holyoak and Mah, 1982; Sadalla *et al.*, 1980; Tversky, 1981,1991a, 1992, 1993, 1996a). Interestingly, these aspects of mental representations of space are reflected in the language people spontaneously use to describe space. That language is qualitative and piecemeal as well (e.g., Denis, 1996; Ehrich and Koster, 1983; Levelt, 1984; Talmy, 1983; Taylor and Tversky, 1996).

MENTAL REPRESENTATIONS OF SPACE FROM EXPERIENCE

It is reasonable to assume that situation models constructed from text are based on the mental representations formed from actual experience with the world. The accumulating evidence for systematic distortions in memory for the spatial world gives some insight into the nature of those mental representations. The view that these are like maps or even schematic maps is giving way to the view that they are more like mixed media, collections of related information, some of which is more visual or spatial, some of which is more linguistic (e.g., Kuipers and Levitt, 1988; Tversky, 1993). Mental representations of environments seem to be put together in an ad hoc way to suit current purposes, rather than being retrieved as unitary wholes. Such mental representations are better characterized as cognitive collages than as cognitive maps (Tversky, 1993). They differ systematically from the two-dimensional Euclidean diagrams that typify maps.

One critical way in which mental representations of environments and environments differ is that mental representations are hierarchical (e. g., Chase, 1983; Hirtle and Jonides, 1985; Maki, 1981; McNamara, 1986, 1992; Stevens and Coupe, 1978; Wilton, 1979). People group cities into states or countries, they group buildings by function. These groupings distort judgments of distance as well as of direction. Distance judgments are also distorted by clutter, that is, the number of entities, for example, buildings or intersections (e.g., Sadalla and Magel, 1980; Sadalla and Staplin, 1980; Thorndyke, 1981), by viewpoint, near or far (Holyoak and Mah, 1982), and by whether the distance is estimated relative to a landmark or to an ordinary building (Sadalla *et al.*, 1980). Direction judgments are distorted by alignment to other proximal elements and by rotation to the horizontal or vertical in a frame of reference (Tversky, 1981). Horizontal and vertical have privileged status in the world as well as in perception (Howard, 1982).

Given these widespread distortions, how can mental representations of the spatial world be characterized? The distortions in distance estimation indicate that mental representations do not directly represent space or distance, but rather represent the entities that exist in space. The distortions in direction estimation indicate that mental representations of directions are not continuous, but rather, tend toward the categorical. Together, these findings suggest that spatial elements are organized relative to each other and to a reference frame, and assimilated toward these. And together, these distortions are unreconcilable in a map-like or Euclidean representation.

MENTAL REPRESENTATIONS OF SPACE FROM DESCRIPTIONS

Spatial Relations

For the most part, research examining spatial mental representations established from description alone has examined simple spatial layouts, with the expectation of finding effects of distance and spatial relations. A number of experiments have demonstrated that people can mentally construct simple scenes composed of familiar objects from description alone (e.g., Clark and Chase, 1972; Glushko and Cooper, 1978; Franklin and Tversky, 1990; Mani and Johnson-Laird, 1982; Taylor and Tversky, 1992; Tversky, 1975). In all these cases, the descriptions specified the spatial relations categorically, using terms like "left", "above" and "in front of." Establishing continuous spatial relations from language alone is complicated by the fact that everyday language is inadequate for expressing continuous spatial relations. Continuous spatial relations can be conveyed by technical language using degrees and meters, but such language is neither produced nor comprehended precisely by ordinary users of language (Flin and Shepherd, 1986; Leibowitz *et al.*, 1993). Morrow and his collaborators have conveyed distance information using a combination of diagrams and text (Morrow *et al.*, 1989; Morrow *et al.*, 1987). However, recent work examining verification times for described routes is more consistent with the view that described distance is represented in gross categories defined by entities, rather than metrically (Franklin, 1989; Rinck *et al.*, 1997).

Spatial Transformations

Mental representations of space derived from descriptions, then, seem to capture approximate but not metric distances and spatial relations, just as mental representations of space derived from experience do. Yet, situation models do not just represent static distances and spatial relations, they are also updated as the described situation changes. That is, they incorporate mental transformations as well as spatial relations. Spatial mental transformations on experienced objects have been extensively studied (e.g., Finke and Shepard, 1986; Shepard and Podgorny, 1978). The mental transformations that people readily perform on actual objects should form the basis for those that can be performed on described objects. This is not to say that everything that people can do in imagination on real objects they can also do on described objects, but rather suggests that transformations that are difficult to do on real objects should also be difficult to do on described ones.

Numerous mental transformations on real objects have been explored, beginning with the now classic experiments on mental rotation (Shepard and Cooper, 1982) and mental scanning (Kosslyn *et al.*, 1978). In addition to mentally rotating and mentally scanning objects, people are also adept at mentally shrinking or enlarging them (Bundesen and Larsen, 1975; Larsen and Bundesen, 1978). In solving geometric analogies, people imagine objects not only rotating, reflecting, enlarging, and shrinking, but also splitting in half, doubling, and changing position (Novick and Tversky, 1987). People are able to imagine their own bodies rotating in space (Parsons, 1987). People can also imagine parts of their bodies, such as their hands and feet, moving in space. In this case, they imagine hands and feet moving in paths that are physically possible (Parsons, 1994). Interestingly, people can imagine objects rotating in space, but are more accurate when the objects are aligned with horizontal and vertical, a result reminiscent of rotation to frame of reference in mental representations of space (Pani *et al.*, 1996; Pani *et al.*, 1995). Other research suggests that mental transformations are not always smooth and analog, as in mental rotation and mental enlargement. When people are queried about the directions of successive pulleys in a system, they appear to mentally animate the system in discrete steps rather than all at once (Hegarty, 1992). Thus, mental transformations seem to be biased in at least two ways: Toward transformations that are possible in the world and toward transformations that are aligned with the horizontal and vertical axes of the world. Moreover, they are not necessarily continuous analog transformations, but in some cases are more categorical, much like the evidence for spatial relations.

To date, the mental transformations in space evoked by language that have been studied have been relatively simple and straightforward. Denis and Cocude (1989) have replicated the studies of Kosslyn *et al.*, (1978) on mental scanning of maps using only descriptions of maps. Just as in the studies of memorized maps, they found increases in scanning time with distance, though the effects were on the whole weaker. Several studies have observed effects of describing a character as moving in a scene (Glenberg *et al.*, 1987; Morrow *et al.*, 1989; Morrow *et al.*, 1987). Other research has found that participants are able to mentally construct images from descriptions of its parts, for example, attaching a capital letter J to the bottom of an upsidedown grapefruit half and realizing that that forms an umbrella (Finke *et al.*, 1989). These simple mental transformations have been readily accomplished by participants. One caveat comes from research inducing mental turning in place from descriptions (Franklin and Tversky, 1990). Although people readily mentally turned in place, this transformation did not have the smooth, continuous character of mental rotation. Turning 180 degrees did not take longer on the whole than turning 90 degrees. Along the same line, there are popular thought problems requiring spatial transformations that seem quite difficult, for example: Think of a sugar cube tilted diamond-like being slowly lowered into a cup of coffee. When it is half-way immersed, what does a cross-section of the cube look like? With so few examples of research, it is difficult to know the constraints of imagination of described spatial transformations. After an introduction to the paradigm, I will describe one modest endeavor in this direction.

Spatial Frameworks

Whether through nature or through nurture or both, people's mental representations of space seem to be derived from their typical experience in it (e.g., Clark, 1973; Fillmore, 1975; Miller and Johnson-Laird, 1976; Shepard and Hurwitz, 1984; Tversky, 1991b). Language, in turn, can draw on, evoke, or reconstruct those mental elements. The world we inhabit is three-dimensional, with an asymmetric vertical dimension determined by gravity and two horizontal dimensions that are unbiased except with respect to a specified viewpoint or origin. Our own bodies, through which we perceive, explore, and experience the world, are more constrained. We have an asymmetric vertical axis defined by head and feet, an asymmetric horizontal axis defined by front and back, and a symmetric left-right axis. Moreover, our experience with the spatial world is not just visual, it is auditory, tactile, kinesthetic and more, so that our mental representations of space are not just multi-modal, but likely to be, at least in part, supra-modal. These facts about the spatial world we perceive and interact with might be expected to constrain mental representations of space and the mental transformations that can be performed on them, and they do.

In order to study people's mental representations of themselves in the world surrounding them, Franklin, Bryant and I have been investigating how people keep track of the objects around them under simple movements, turning in place, lying down, and standing up. Our initial research used narratives to convey these worlds, and we found that people had no trouble constructing mental representations of them as well as updating them as the described situation changed (Franklin and Tversky, 1990). The narratives described you, the reader, in a three-dimensional environment, such as an opera house, museum, or hotel lobby, surrounded by objects particular to the environments located beyond your head, feet, front, back, left, or right. Once participants had studied and learned the environments, they turned to a computer that redirected them to face another object in the scene, and then queried them for the current locations of the other objects with successive direction probes, randomly ordered, left, right, front, back, head, and feet. Performance was highly accurate; the data of interest were the reaction times to the various directions.

We considered several models to account for the times to retrieve objects in the probed directions about the body. Because the environments did not favor any direction over any other, the *Equiavailability Model* would predict equal reaction times to all directions (Levine *et al.*, 1982). The data did not corroborate this account, nor did they support an account based on classic theories of imagery (e.g., Kosslyn, 1980; Shepard and Podgorny, 1978). According to the latter account, participants would imagine themselves in the described situations, facing forward. To ascertain which object was lying in the probed direction, they would imagine themselves turning to search that direction. The *Mental Transformation* account would predict fastest reaction times to front, next fastest to directions 90 degrees from front, namely, head, feet, left, and right, and slowest times to back, 180 degrees removed from front.

Although the Equiavailability and Mental Transformation accounts did not fit the data, the data did conform to the *Spatial Framework* analysis. According to Spatial Framework analysis, participants construct a mental spatial framework out of the three body axes and associate objects to them. The axes differ in accessibility, depending on asymmetries of the body and of the world. For an upright observer, the head/feet axis should be most readily accessed because of its asymmetries and because it is correlated with the only asymmetric axis of the world, that defined by gravity. Front/back should be next, due to its asymmetries, and left/right, lacking prominent asymmetries, should be slowest. This pattern of reaction times has been found in dozens of experiments (e.g., Bryant *et al.*, 1992; Franklin and Tversky, 1990; Franklin *et al.*, 1992). Further support for the model comes from the case where the observer in the scene reclines and turns from front to side to back. In this case, no axis of the body is correlated with gravity, so locating objects depends only on the relationship to the body. In this case, we argued that the front/back axis should be fastest as the front/back asymmetries in perception and behavior are especially important; that axis separates the world that can be observed and manipulated from the world that cannot. In fact, for the reclining observer, times to locate objects to front and back are faster than times to locate objects to head and feet (Franklin and Tversky, 1990).

Since the initial studies, we have gone on to explore a number of variations of the original situation. Variants of the spatial framework analysis account for patterns of retrieval times to probed directions for variants in the situation, including points of view external to the scene as well as internal (Bryant *et al.*, 1992) and scenes with more than one viewpoint (Franklin *et al.*, 1992). In yet another variant, participants learned the scenes from observing them rather than from narrative, and responded from memory just as participants in the experiments where scenes were described had done (Bryant *et al.*, in press). The experiments reported here investigate yet another extension, one where instead of describing the person in the scene as turning in place and the surrounding environment as stationary, narratives describe the person as stationary and the surrounding environment as turning.

There are reasons for expecting that the case of imagining spatial transformations of an environment would be more difficult that the case of imagining spatial transformations of one's own body. Except for earthquakes, in nature, it is people who move in environments, not environments that move around people. Of course there are haunted houses and other such dubious pleasures in amusement parks, but they are atypical and recent inventions, accounting for their attraction. Train and automobile travel can give the impression of being stationary while the world passes by but these are also recent inventions in our evolutionary history, though a small élite has probably been carried going farther back in our history. Even so, classic research on kittens carried in baskets yoked to kittens who performed movements has shown that the carried kittens learned little compared to the mobile kittens (Held and Hein, 1963). Formally, however, the situation where the person is turning in a stationary environment and the situation where the environment is turning around a stationary person are identical, so that a propositional account of behavior would predict that there should be no differences between the two cases.

Experiment 1: Turning Person vs. Turning Environment

The narratives for this experiment described the participants as floating in a weightless space house designed by NASA, visiting different rooms. Movement could be accomplished either by describing the participant as moving or by describing the room as moving around the participant (Kim, 1992). In what turned out to be a pilot experiment, two types of movement were described, turning on the same plane, or rotating from plane to plane. Although participants in the moving person condition had no difficulty carrying out this experiment, participants in the moving room condition were unable to do so. They quickly got disoriented. Many of them realized that imagining themselves turning or rotating in one direction was equivalent to imagining the room turning or rotating in the opposite direction, but they said they were still unable to perform the rotations. Therefore, we carried out an experiment where participants were described as only turning on a horizontal plane and not also rotating to another plane.

There were two sets of reaction times of interest: The times to reorient to a new facing, and, after the reorientation, the times to retrieve objects at the probed directions. The expectation was that for this conceptually easier situation, participants would be able to perform the task, but that reorientations would be more difficult, that is, take more time, when the room was described as moving rather than the observer. Once the reorientations had been accomplished, it was expected that the spatial framework pattern of times to retrieve objects in the probed directions would emerge for both cases.

Method

Participants. Twenty-four Stanford undergraduates participated either for cash or for credit in an introductory psychology class. Six males and six females participated in the moving person condition and five males and seven females participated in the moving room condition.

Procedure and descriptions. Participants first read and signed a consent form. Then they were given general instructions about the nature of the experiment. Following this, participants in the moving person condition read the following general description of the situation:

"You are visiting a large space museum created by NASA. The purpose of this museum is to show people what it would be like to live in outer space. The largest exhibition in the museum is the newly built "spacehouse", a full-scale model house created as a prototype that may eventually be used to colonize the moon. It looks like a normal contemporary style house. However, there is very little gravity in the house, and most of the objects are bolted to the floors, ceilings, and walls to keep them securely in place. You enter the house and find yourself floating in mid-air, a few feet above the floor. You are in an upright position, *but discover that you can change the position of your body by shifting your weight in the desired direction. You can rotate your body in three ways. You can rotate your body to the left 90 degrees, so that you face the object that was previously to your left, or you can rotate to your right 90 degrees, so that you face the object that was previously to your right. You can also rotate your body around 180 degrees so you face the object that was previously at your back (behind you).* At all times, your body is in a straight, upright position. The most striking feature you notice is that objects are not where you are accustomed to seeing them. Some objects are even attached to the ceiling. This seems appropriate since the lack of gravity allows you to move equally well (or with equal difficulty) in all directions. You decide to explore the house."

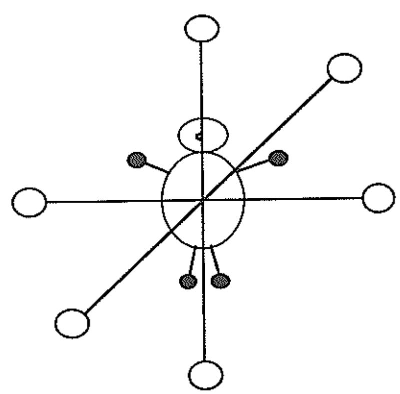

Figure 1

Participants in the moving room condition read the same general description, with the section below substituting for the section in italics above:

... and it is difficult to move. In the house is a small hand-held device which looks like a remote control. It has several buttons on it. The buttons are labeled "left", "right" and "180 degrees." You discover that the rooms in the spacehouse rotate, so that you do not need to change positions, and that the hand- held device controls the rotation of the room. For instance, if you press the "left" button, the room rotates to the left 90 degrees, so that the wall which used to be in front of you is now to your left. Pressing the "right" button causes the room to rotate 90 degrees to the right, so that the wall that was previously in front of you is now to your right. If you press the "180 degrees" button, the room rotates 180 degrees, so that the wall which used to be in front of you is behind you."

Following this, participants read descriptions of visits to each room, preceded by a list of the objects in that room. For example, the living room contained a lamp, a newspaper, a plant, a portrait, a stereo, and a television. The objects for each of the six rooms were selected to be common, distinctive, appropriate to the setting, and approximately the same size. They were described as being beyond the observer's head, feet, front, back, left, and right, with particular positions determined randomly. A diagram of the situation appears in Figure 1. The description of the living room located the objects relative to the observer as follows:

"You are in the living room of the spacehouse. The living room is spacious, and the ceiling is made of glass, giving it the appearance of a greenhouse. You are in an upright position, suspended several feet above the floor. Above your head hanging from the ceiling is a plant. It is plastic, and is full of leaves. Below your feet on the ground is a lamp. It opens up at the top, giving it the appearance of a torch. In front of you, built into the wall is a television. It is turned on and a soap opera is playing. To your right, there is a newspaper attached to the wall. The cover story is about a civil war that has broken out on Mars. You turn to your left and there is a stereo on top of a shelf which is attached to the wall. It is made up of several components. You look behind you and see a portrait on the wall. It pictures a family with two children and a dog."

Participants were allowed to study each story as long as they wished, until they felt they knew the locations of the surrounding objects. Then they turned to an IBM-AT computer for the continuation. Participants read sentences on the screen one at a time, striking a key to read the next sentence.

The first sentence reoriented the participant by describing either the participant or the room as rotating in one direction or another. When participants judged that they had been reoriented, they pressed a key to continue the experiment. The time from presentation of the reorientation sentence to pressing the key was the reorientation time. This was followed by a statement describing the object currently faced, and then a filler sentence not related to the object now faced to eliminate any priming effects for that object. Following this, one of the six direction probes appeared in random order: head, feet, front, back, left or right. Participants were instructed to press any key as soon as they knew which object was in that directions. The time between the direction probe and the key press was called RT1. After the key press, a list of the six objects appeared on the screen, and participants pressed a key corresponding to the number of the objects. This was called RT2. In most of the experiments run so far, RT1 reflects the times to respond to each of the directions and to determine the appropriate object whereas RT2 reflects only the time to find the name of the