

Mental Representations of Perspective and Spatial Relations From Diagrams and Models

David J. Bryant
Humansystems

Barbara Tversky
Stanford University

In previous research (D. J. Bryant, B. Tversky, & N. Franklin, 1992; N. Franklin & B. Tversky, 1990), the authors showed that spatial knowledge conveyed by descriptions and direct experience induces participants to take the perspective of a character surrounded by objects. In this study, the authors used models and diagrams to convey the same information. With models, as with descriptions and experience, participants adopted the character's perspective (the *spatial framework* analysis). With diagrams, participants took an outside perspective (the *intrinsic computation* analysis). Even when informationally equivalent, different depictions made salient different aspects of the world. When instructed, however, participants were able to take either the inside or the outside perspective in memory for both diagrams and models. Depth cues in depictions also govern participants' perspective. When diagrams contained rich pictorial depth cues, participants used the spatial framework analysis, and when models were viewed without access to depth cues, participants relied on the intrinsic computation analysis.

People's knowledge of the world comes not only directly, from experiencing the world, but also indirectly, from descriptions and depictions of the world. Perhaps because of its significance, spatial knowledge has been conveyed by external representations since prehistory. Maps, whether from stone, clay, wood, bark, or paper, have been invented by many cultures (e.g., Brown, 1949; Wilford, 1981). Spatial language alone can act like a map, effectively conveying spatial relations and relative distances (e.g., Bryant, Tversky, & Franklin, 1992; Denis & Cocude, 1989; Franklin & Tversky, 1990; Glenberg, Meyer, & Lindem, 1987; Mani & Johnson-Laird, 1982; Morrow, Greenspan, & Bower, 1987; Taylor & Tversky, 1992) and allowing updating of relative positions and perspectives as new information becomes available (e.g., Bryant et al., 1992; Franklin & Tversky, 1990; Franklin, Tversky, & Coon, 1992; Glenberg et al., 1987; Morrow et al., 1987). Language describing space is so fundamental that it is used to express other, nonspatial concepts, such as time, mood, and power (e.g., Clark, 1973; Lakoff & Johnson, 1980).

External Representations: Diagrams and Models

Diagrams and models are external graphic representations or depictions that consist of elements and the spatial relations among them (Tversky, 1995a). As such, they are external stimuli with their own spatial properties. In particular, relations in the represented world are mapped onto spatial relations in the graphic representation (Tversky, 1993, 1995a; Tversky, Kugelmass, & Winter, 1991). For example, in corporate organization diagrams, the vertical spatial relations represent power. In the case of the depictions studied here, spatial relations in the represented world are mapped onto spatial relations in the diagrams and models.

Depictions, however, schematize the situations they represent and require interpretation. What is schematized, and how it is schematized, can affect how a depiction is interpreted and used. The third dimension, for example, is an important factor that must be schematized in depictions of spatial situations. As we use the terms, *models* convey all three spatial dimensions directly; *diagrams*, by contrast, may depict three-dimensional (3D) relations but are themselves two dimensional (2D). They may use a number of conventions for conveying depth, including relative size, occlusion, height in the picture plane, and converging lines, but they necessarily lack binocular cues. Also, diagrams more often than models use verbal and symbolic information to convey spatial information. Thus, diagrams convey the 3D structure of an environment indirectly, whereas 3D models convey that information directly.

One goal of the present research is to explore how the differences between diagrams and models lead people to create different kinds of mental models of depicted environments. A second goal is to determine whether diagrams and models necessarily induce one kind of mental representation or whether individuals can alter their representations on the basis of instructions. A third goal is to explore what features

David J. Bryant, Humansystems, Inc., Guelph, Ontario, Canada; Barbara Tversky, Department of Psychology, Stanford University.

This research was conducted at Northeastern University and sponsored by a Research and Scholarship Development Fund grant from Northeastern University and by the Air Force Office of Scientific Research, Air Force Systems Command, U.S. Air Force, under grant or cooperative number AFOSR 94-0220. We are grateful to Sophia DeRosa, Carol Gapsky, Marnie Nilsson, and Virginia Ricker for their assistance in conducting this research and to Maggie Lanca, Gordon Logan, Ruth Maki, three anonymous reviewers, and the Language and Cognition Group at Northeastern University for their helpful comments.

Correspondence concerning this article should be addressed to David J. Bryant, Humansystems, Inc., 111 Farquhar Street, Second Floor, Guelph, Ontario, Canada N1H 3N4. Electronic mail may be sent to dbryant@humansys.com.

of diagrams and models, especially those related to conveying depth, lead to differences in mental representation.

Mental Representations of a Paradigmatic Spatial Situation

Paradigmatic Scene

One spatial situation people carry around with them at all times is their own bodies surrounded by the objects in their immediate environment. This is a useful situation for studying people's spatial mental models, because people appear to keep track of objects effortlessly as they move about the world. In our studies, we investigated participants' understanding of depictions of a scene in which a character is situated in a natural setting, such as a kitchen or living room, with objects, such as a pot or spoon, located to the six body sides (head, feet, front, back, left, and right). The situation is illustrated in Figure 1, which is also an example of one of the diagrams used in our experiments. After learning a scene, participants were informed that the character had moved to face a new object, and participants were then probed for the objects lying in the six directions from the character's body.

Because certain body axes have a favored status in people's interactions with the world, they are more salient to thinking about spatial relations, and this leads to differences

in retrieval times for spatial relations. Which axes are most salient depends on the mental perspective one adopts. There are at least two perspectives that one could effectively use for our paradigmatic scene, the *inside* and *outside* perspectives. Which perspective a participant adopts will determine his or her mental representation and how he or she accesses spatial directions, which will in turn affect retrieval of spatial information.

Spatial Framework Analysis

To represent the scene, participants could adopt the inside perspective of the character. This perspective leads to a particular organization of objects in participants' mental models of scenes. Franklin and Tversky (1990) developed the *spatial framework analysis* (based in part on previous analyses of spatial language and cognition by Clark, 1973; Fillmore, 1976; Levelt, 1984; Miller & Johnson-Laird, 1976; and Shepard & Hurwitz, 1984; among others) to explain patterns of retrieval times of spatial information from narrative descriptions of scenes. According to this analysis, people mentally place themselves in the place of the character and use their own head/feet, front/back, and left/right axes to code locations of objects. In other words, they apply an egocentric frame of reference.

The accessibility of objects in memory in this analysis

Bob in the Kitchen

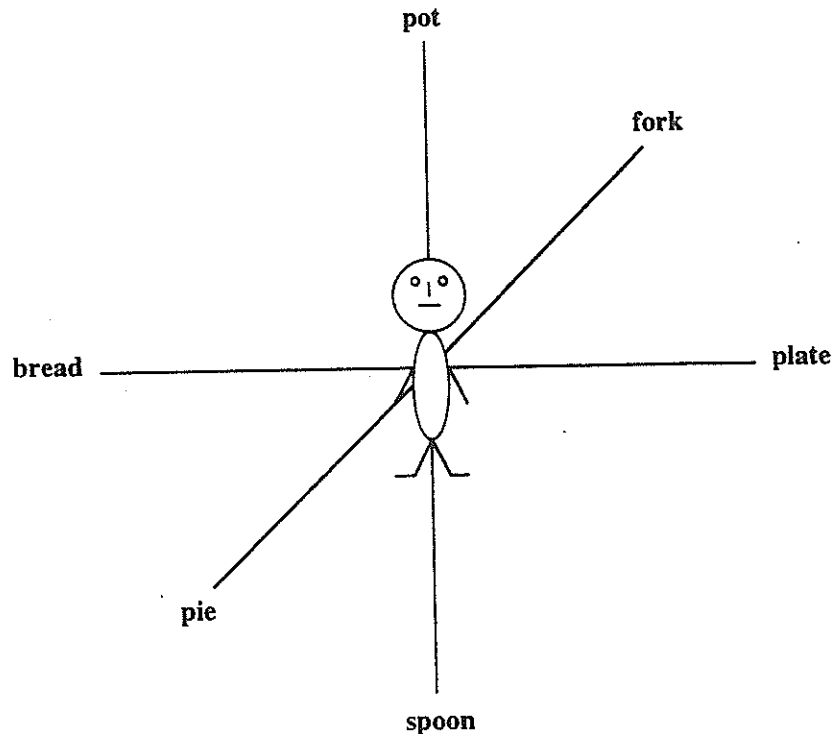


Figure 1. "Bob in the Kitchen": Paradigmatic scene and example of a diagram used in Experiments 1-3. Depth is indicated by the diagonal line; the pie to the lower left projects out of the page and the fork to the upper right projects into the page.

depends on the characteristics of the participant's own body and perception of space, rather than those of the character. For the body, the front/back and head/feet axes are more salient than the left/right axis because of their biological asymmetries. Because the perceptual and motor apparatus are oriented forward, the front/back axis separates the world that can be easily perceived and manipulated from the world that cannot. This gives that axis a slight advantage over the head/feet axis, but one that can be countered by the asymmetries of the physical world. The world has one salient asymmetric axis: the vertical axis of gravity. Thus, when the character in the scene is described as upright with the body's natural axis of rotation aligned with gravity, the asymmetries of the body and world combine to make the head/feet axis more accessible than the front/back axis. Thus, participants should be faster to identify objects to the head/feet than front/back, and faster to identify objects to the front/back than left/right.

When the character in the scene is described as reclining, however, and turning from back to side to front, the axis of gravity no longer corresponds to a body axis, so the front/back body axis is most accessible. In this case, the biological asymmetries of front/back, which are more extreme than those of head/feet, predominate. Thus, participants should be faster to respond to objects to front/back than head/feet, and slowest to objects to left/right. These predictions have been upheld in more than a dozen experiments (Bryant, Lanca, & Tversky, 1995; Bryant & Tversky, 1992; Bryant et al., 1992; Bryant, Tversky, & Lanca, 1998; Franklin & Tversky, 1990; Franklin et al., 1992).¹

Participants construct similar mental models from experiencing a situation as from reading about a situation. When participants in one experiment learned an environment by standing in it themselves and responded to direction probes from memory, their patterns of reaction times (RTs) to retrieve objects were the same as when learning was from description (Bryant et al., 1998).

Intrinsic Computation Analysis

A second way that participants could represent scenes is by taking the outside perspective and imagining themselves looking at the entire scene from some external position. This perspective affords a view of the character and objects. Given the outside vantage point, participants do not rely on their own body axes to code space but instead use a frame of reference centered on an external object or person. In our paradigm, participants locate objects by determining where the character's body sides are and naming the objects associated with each side. Accessibility of objects depends on the cognitive and perceptual mechanisms used to identify the sides of the character.

We call this procedure the *intrinsic computation analysis*, and it can be invoked for any object that has acknowledged intrinsic sides, front, back, top, bottom, and consequently left and right, such as people and cars, but not balls or trees (e.g., Fillmore, 1975; Levelt, 1984; Levinson, 1996; Miller & Johnson-Laird, 1976).² Some intrinsic sides are more

readily determined than others. Research indicates that people first determine the top/bottom axis of an object, which is the head/feet axis of a person (Braine, Plastow, & Greene, 1987; Jolicoeur, 1985; Maki, 1986; Rock, 1973). In particular, Rock showed that to identify what an object is, one needs to know how it is oriented (i.e., where its top is). The front/back axis of an object must be determined prior to the left/right axis as the left/right axis can only be defined with respect to the top/bottom and front/back axes. Corroborating this analysis, people are faster to identify the tops and bottoms of objects than the fronts and backs of objects at all orientations (Jolicoeur, Ingleton, Bartram, & Booth, 1993). They are also faster at identifying asterisks at the top and bottom than at the left and right for all orientations (Corballis & Cullen, 1986). Identification of sides in the intrinsic frame does not depend on the orientation of the object or viewer. Thus, participants using the intrinsic computation analysis should always be fastest to identify objects at the head/feet, then the front/back, and finally the left/right of the person in the scene, irrespective of orientation.

The intrinsic computation pattern has been observed when participants respond to direction probes while viewing 2D diagrams. Logan (1995) presented diagrams of 2D slices of our 3D situation, consisting of schematic heads in front, profile, or top views, with colored dots located in the appropriate directions. The heads were presented upright or rotated 90°, 180°, or 270°. The participant's task was to make judgments about the directions of colored dots from the heads. Logan's data fit the intrinsic computation pattern of response times, head/feet fastest followed by front/back followed by left/right, for all orientations of the head. His data did not fit the spatial framework pattern. Although he referred to the spatial framework analysis of Franklin and Tversky (1990) to explain his data, Logan did not note the inconsistency of his data with theirs for the cases that were not upright. Because Logan did not recognize the inconsistency, his explanation of his data is inadequate and does not

¹ The reclining situation, in which gravity and verticality no longer correlate with any body axis, eliminates two alternative explanations for the primacy of head/feet in the upright situation (see Bryant et al., 1992; Franklin & Tversky, 1990). The primacy cannot be due to the fact that the head/feet axis happened to be the vertical axis for the upright orientation. In the reclining case, left/right corresponded to the vertical axis for half of the trials, but reaction times (RTs) were slower than those to head/feet and front/back in the horizontal plane. The primacy also cannot be attributed to the fact that objects to the head and feet were constant with rotations of the character. In the reclining case, objects to head and feet were still constant with rotations, but RTs to head/feet were slower than those to front/back when objects changed with each rotation.

² The intrinsic computation analysis is not the same as the external spatial framework described in previous research (Bryant et al., 1992). Both assume a mental perspective outside or external to the scene. In the external spatial framework, however, participants locate objects with respect to their own body sides. Participants using intrinsic computation locate objects in relation to the intrinsic sides of the character in the scene.

account for the inconsistency of his results with those of Franklin and Tversky. The intrinsic computation analysis, however, does provide an adequate explanation for Logan's data and indicates that his participants used an outside rather than inside perspective.

The spatial framework and intrinsic computation analyses can easily be distinguished by their predictions for a character who is reclining. According to the spatial framework analysis, participants mentally place themselves in the position of the character in the scene. When the character is reclining, the head/feet axis is out of its canonical alignment with gravity, and participants are faster for front/behind than head/feet relations. According to the intrinsic computation analysis, the participant identifies the sides of the character from an outside perspective, beginning with the axis of the intrinsic top. As a consequence, participants should be faster to head/feet than front/back at all orientations of the person.

Present Research

Selecting a physical model of the paradigm situation for the present research was not difficult, as a model is inherently 3D. The model consisted of a doll with schematic pictures of objects suspended in the appropriate positions. Selecting a diagram was more difficult. A diagram is necessarily 2D, but in this case, it needed to convey a 3D situation. In Western art, the third dimension has been expressed by convergent perspective since the Renaissance. However, this convention has not been universally adopted (e.g., Asian art) because it can distort shape and size. Rather than using a biased, complex, and cluttered converging perspective drawing to convey the scene, we schematized the depth dimension by adapting another common convention for conveying depth—a diagonal line—and eliminated irrelevant detail. This convention is used spontaneously by children (Braine, Schauble, Kugelmass, & Winter, 1993) and is standard in Chinese art (Willats, 1990).

In the first experiment, we examined mental representations spontaneously established from diagrams and models. In the second experiment, we examined the effects of instructions to interpret the diagrams and models on the mental representations established. In the final two experiments, we examined the characteristics of diagrams and models that induce the use of intrinsic computation or spatial framework analyses. Because models convey 3D information directly, but diagrams do not, we expected models to induce taking the perspective of the character, yielding the spatial framework pattern of RTs, and diagrams to induce taking the outside perspective, yielding the intrinsic computation pattern.

The current experiments consider the upside-down orientation, which has not been examined in previous studies (e.g., Bryant et al., 1992; Franklin & Tversky, 1990). Predictions of the intrinsic computation analysis are the same as for all orientations, and predictions of the spatial framework analysis match the upright orientation. The head/feet axis is again aligned with gravity, but in a

noncanonical orientation. The asymmetries of front/back could render this axis most salient, as it does for the reclining posture. People, however, naturally rotate around the head/feet axis as they navigate the world, as when a person turns to walk in another direction. In the current paradigm, too, the character rotates around the head/feet axis. Having the natural axis of rotation aligned with gravity, even in the opposite direction, should render it more salient than front/back. Thus, for an upside-down character, participants should show fastest access to head/feet, followed by front/back, followed by left/right. Participants, however, should be slower overall because of the character's noncanonical orientation.

General Method

All experiments followed the same general method. Changes in materials and procedures specific to each experimental manipulation are discussed separately for each experiment. The materials and manipulations are summarized in Table 1, with predictions for each condition.

Participants

Participants in all experiments were Northeastern University undergraduates with normal or corrected-to-normal vision who participated for credit in an introductory psychology class. The numbers of participants in each experiment are listed in Table 1.

Scenes

Diagram Condition

Participants learned four critical scenes, indicated in Table 2, and one for practice. Each scene depicted a setting with a character surrounded by six objects (see Figure 1). The name of the character and the type of setting were printed at the top of the diagram. In half the scenes, the name given to the character was female and in the other half it was male. The settings and the objects were selected to be familiar and common and to form a coherent scene. The locations of objects were selected randomly. In all diagrams, the character was shown facing forward.

The character was 3.9 cm (1.5 in.) long. The vertical and horizontal axes were 11.7 cm (4.5 in.) long, and the diagonal axis was slightly shorter (approximately 11.05 cm, or 4.25 in.). In all diagrams, the diagonal was drawn from the lower left to the upper right of the page. Braine et al. (1993) observed an early tendency in children to interpret objects to the left and lower in pictures as being closer than objects to the right and higher. This suggests a bias to interpret the left end-point of a diagonal as nearer than the right end-point. The name of an object was printed at the end of each axis.

Model Condition

Participants learned four critical scenes plus one for practice. Four of the scenes (bedroom, construction site, kitchen, and living room) were the same as those of the diagram condition. The other (backyard) was adapted from materials used by Bryant et al. (1998). A "Homer Simpson" doll (28 cm tall) was placed in the center of the model. The doll stood on a platform 14 cm high and

Table 1
Summary of Experimental Conditions

Experiment and n^a	Study material	Experimental manipulation ^b	Predicted mental representation
Experiment 1 ^c			
32 ^d (13/19)	Diagram		Intrinsic computation
24 (12/12)	Model		Spatial framework
Experiment 2 ^c			
20 ^e (10/10)	Diagram	Inside perspective instructions	Spatial framework
20 (11/9)	Model	Outside perspective instructions	Intrinsic computation
Experiment 3 ^f			
16 (8/8)	Standard diagram	No depth cues	Intrinsic computation
16 (8/8)	Intermediate diagram	Converging lines	Spatial framework
16 (8/8)	Perspective diagram	Converging lines, relative size, texture gradient	Spatial framework
Experiment 4 ^f			
16 (8/8)	Standard model	Normal depth cues	Spatial framework
16 (8/8)	Impoverished model	Minimal depth cues	Intrinsic computation

^aNumbers of male and female participants are indicated, respectively, in parentheses. ^bThe orientation of the character and the objects' directions relative to the character were manipulated within-participant in all experiments. This column lists additional manipulations. ^cBetween-participants design. ^dData of 1 participant were discarded due to high error rate. ^eData of 3 participants were discarded due to high error rate. ^fWithin-participant design.

could be rotated to face four directions or reclined to face in four directions. Drawings of the objects in each scene were hung from narrow wooden shafts to the front, back, head, feet, left, and right of the doll, such that they faced the participant at all times.

Procedure

Diagram Condition

Participants received detailed instructions about the procedure before beginning. Participants were instructed that the diagonal axis represented an axis in depth and were asked to think of the lower end as projecting out of the page toward them. Participants were told that the diagrams conveyed 3D environments and that they should attempt to think of what that setting would be like. They were encouraged to think of the character as standing or reclining on some kind of platform in the center of the scene and to elaborate the setting to help them remember the scene. They were instructed to study each diagrammed scene, taking care to learn the names of the objects and where they were located. Participants were allowed to study the diagram for as long as they wished, then returned it to the experimenter. They then proceeded to the direction probes, which were presented by computer.

A block of direction probes always began with a sentence presented on the computer screen telling the participant that the character had turned to face another object and/or changed orientation (from upright to reclining, reclining to upside down, etc.). The sentence specified the direction in which the character

turned and stated explicitly the character's resultant orientation and which object the character now faced. When participants understood this sentence, they pressed the space bar of the computer keyboard to receive a series of six direction probes. Probes consisted of the names of the six body directions in relation to the character (front, back, head, feet, left, and right). Participants were specifically instructed not to interpret the probes in relation to themselves. In response to a probe, participants pressed the space bar as soon as they knew which object was located at that direction, without sacrificing accuracy. The time participants took to do this was the critical RT. After participants pressed the space bar, the names of the six objects appeared in a line on the screen in random order, numbered 1 to 6. Participants pressed a numbered key corresponding to the correct object as quickly and accurately as possible. This served as an accuracy check. Direction probes were separated by 500 ms of blank screen, and a series continued until all six directions had been probed.

The character changed orientation four times during the probing procedure: upright, reclining with the head pointed left (in the picture plane), upside down, and reclining with the head pointed right. Changes in orientation proceeded counterclockwise in the picture plane in all scenes. After participants completed the first block of six probes, the character was rotated around the head/feet axis in that same orientation in three subsequent blocks of trials. The character rotated counterclockwise in all scenes. After four blocks in one orientation, the participant was told that the character had changed orientation and completed four rotations in that

Table 2
Scenes and Objects Used in Experiments 1 and 2

Scene	Character	Objects
Construction site	Harry	axe, bucket, jackhammer, ladder, shovel, wheelbarrow
Barn	Nancy	brush, hay, lantern, pail, saddle, shears
Bedroom	Steve	dress, hat, pants, purse, shirt, sock
Kitchen	Bob	bread, fork, pie, plate, pot, spoon
Living room	Sally	chair, clock, lamp, painting, table, vase

posture, and so on. Participants completed 16 blocks of probes for each scene.

Model Condition

The procedure was the same as that of the diagram condition, except that participants studied a physical model of scenes, rather than a diagram, and the character in the scene was always Homer Simpson.

Design

In all experiments, within-participant independent variables included direction (front, back, head, feet, left, and right) and orientation (upright, upside down, reclining to the left, and reclining to the right). Manipulations specific to each experiment are listed in Table 1. The critical dependent variable in all experiments was RT. Orders of presentation of scenes, of orientation within scenes, of rotation within block of probes, and of direction probes were counterbalanced in the same way for all experimental conditions. Equal numbers of participants were assigned to eight random orders of presentation of the four scenes. Four versions of the block sequence were constructed for each scene. In one version, the character began upright and rotated counterclockwise across blocks of probes. In other versions, the character began reclining to the left, upside down, and reclining to the right. Version was counterbalanced across participants such that each participant received one scene in which the character began the probing procedure in each orientation. Within a scene, the character rotated around its head/feet axis in a clockwise fashion, and in the other half counterclockwise. Direction probes within a block were assigned one of six counterbalanced orders that assured that each probe appeared in each serial position an equal number of times.

Experiment 1: Locating Objects From Memory of Diagrams and Models

This experiment documents differences in representations of diagrams and models. Because of the strong cues to depth, participants who learn scenes from a model should adopt the inside perspective of the character in the scene and use spatial frameworks. Depth cues convey a detailed 3D environment, making it easy for participants to mentally place themselves in the scenes. Because of the weak depth cues and small size of diagrams, participants who learn scenes by diagram should adopt the outside perspective and use intrinsic computation.

Results

In this and all subsequent experiments, a probability criterion of .05 was assumed for statistical tests, unless otherwise stated.

Data Treatment

Diagram condition. Participants made errors on 7.4% of probes, and these data points were discarded from analysis. Outliers, defined as RTs greater than a participant's depiction by orientation by direction cell mean plus two standard deviations, accounted for 5.1% of the data and were also

discarded. In addition, all RTs from a total of three scenes from 3 participants were discarded because participants made more than 16 errors in these scenes. The data of 1 participant were discarded in its entirety because the participant averaged more than 16 errors per scene. Outliers in this and all conditions of all experiments were generally equally distributed across direction and orientation conditions. In particular, there were no more outliers to the relatively difficult left and right probes than other directions.

Men and women generally displayed the same patterns of RTs. There was no overall effect of participant gender, $F(1, 30) = 2.56$, $MSE = 51.70$, but the three-way interaction of gender, orientation, and direction was significant, $F(15, 450) = 2.25$, $MSE = 1.08$. This finding seems to reflect that men were slightly faster overall for the upright orientation, but not for others, and that women tended to show extreme RTs to right and left probes for reclining orientations. The effect of gender in the diagram condition was considered in the analysis of response patterns for each orientation.

Model condition. Participants made errors on 5.1% of probes, and 5.1% of the data were outliers. Men and women displayed the same patterns of RTs. There was no overall effect of participant gender, $F(1, 21) = 0.03$, $MSE = 0.58$, and this factor did not interact with any other.

Effect of Type of Depiction

RT data were subjected to an analysis of variance (ANOVA) with type of depiction as a between-participants variable and orientation and direction as within-participant variables (see Table 3). RTs did not differ overall between the diagram and model conditions, but different patterns of RTs were observed in the two conditions. For this reason, the effects of orientation and direction were examined in separate ANOVAs for the diagram and model conditions. The results of these ANOVAs are also shown in Table 3.

Diagram Condition

Mean RTs are presented in Table 4. The pattern of RTs for reclining to the right did not differ from the pattern for reclining to the left, and these two conditions were collapsed to a single reclining condition. Similarly, in all subsequent experiments, data were collapsed to form a single reclining condition because there were no significant differences between reclining to the right and reclining to the left.

Because the diagrams contained few depth cues, participants should have adopted an outside perspective and used intrinsic computation to locate objects. Thus, for all orientations, participants should have been faster to identify objects to the head/feet than front/back, and slowest to identify objects to left/right. Participants should also be slower overall for nonupright orientations because it is more difficult to identify the sides of the character when it is reclining or upside down.

Direction significantly affected RTs in the diagram condition (see Table 3). When the character was upright, participants responded faster to head/feet than front/back,