Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Cognitive Psychology 51 (2005) 285–333

Cognitive Psychology

www.elsevier.com/locate/cogpsych

Acquiring new spatial intuitions: Learning to reason about rotations ☆

John R. Pani *, Julia H. Chariker, Thomas E. Dawson, Nathan Johnson

Department of Psychological and Brain Sciences, University of Louisville, Louisville, KY 40292, USA

Accepted 12 June 2005
Available online 13 September 2005

Abstract

There are certain simple rotations of objects that most people cannot reason about accurately. Reliable gaps in the understanding of a fundamental physical domain raise the question of how learning to reason in that domain might proceed. Using virtual reality techniques, this project investigated the nature of learning to reason across the domain of simple rotations. Learning consisted of the acquisition of spatial intuitions: there was encoding of useful spatiotemporal information in specific problem types and a gradual accumulation of this understanding across the domain. This pattern of learning through the accumulation of intuitions is especially interesting for rotational motion, in which an elegant domain-wide kinematics is available to support insightful learning. Individual ability to reason about rotations correlated highly with mastery motivation, skill in fluid reasoning, and skill in reasoning about spatial transformations. Thus, general cognitive advantages aided the understanding of individual rotations without guaranteeing immediate generalization across the domain.

☆ The first and second authors were supported by a grant from the National Library of Medicine, National Institutes of Health, during preparation of this manuscript (Grant 1 R01 LM008323-01A1; Histological Reasoning: Visual Cognition in Microanatomy). The second, third, and fourth authors received support from the Grawemeyer Awards during conduct of this research. We thank Carolyn Mervis for advice and training in the use of psychometric tests.

* Corresponding author.
E-mail address: jrpani@louisville.edu (J.R. Pani).
URL: http://www.louisville.edu/~jrpani01/ (J.R. Pani).

0010-0285/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.cogpsych.2005.06.002

Keywords: Learning; Visual cognition; Reasoning; Mental models; Qualitative reasoning; Perceptual organization; Spatial organization; Mental rotation; Virtual reality

1. Introduction

Rotational motion is fundamental in mathematics and physical science, common among ecological motions, and a mainstay of engineering. The ability to predict the outcome of a rotation is a prototypic instance of physical reasoning. Thus, people can reason about the rotations of a great variety of objects, the spatial properties of rotation are simple and elegant, and formal systems for describing rotation reduce it to a few fundamental principles. Indeed, it has sometimes seemed that the ability to think about rotations is a systematic internalization of constraints on the physical world (e.g., Cooper & Shepard, 1973; Shepard, 2001; Shepard & Metzler, 1971).

These characteristics of cognition with regard to rotation make it especially interesting that there are elementary instances of rotational motion that most people are very poor at reasoning about (e.g., Just & Carpenter, 1985; Massironi & Luccio, 1989; Pani, 1989, 1993, 1997; Pani & Dupree, 1994; Parsons, 1987, 1995). Such constraints on spatial knowledge are most apparent when experimental tasks are not confined to the classical paradigm of mental rotation (i.e., when the test of knowledge is not a discrimination of an object and its mirror reversal; see Pani, 1993; Parsons, 1995). For example, suppose an individual is shown a physical assembly like that illustrated in Fig. 1 and is asked to indicate which direction the dish would be facing after a rotation around the long shaft (e.g., 180°). For the case at the left (Fig. 1A), the person would respond fairly quickly and the answer likely would be

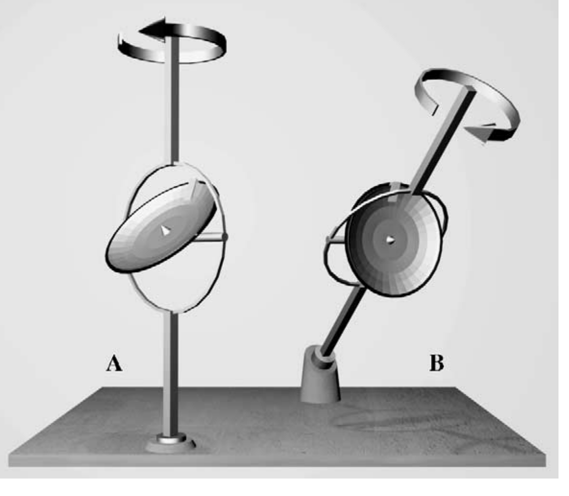

Fig. 1. Two configurations in which the main shaft can be an axis of rotation for the dish assembly.

correct. For the case at the right, the person probably would take more than a minute to consider the problem. They would then give an answer that was incorrect by 45° or more.

Reliable gaps in the ability to reason in a fundamental domain of physical transformation raise the question of how learning to reason in that domain might proceed. And an understanding of the progress of learning can inform theories of how reasoning takes place in each instance. This paper pursues the question of how learning progresses when people learn to reason accurately across the entire domain of simple rotations (i.e., rotations about single fixed axes).

The remainder of this introduction is divided into four parts. The first part explains why certain rotational motions are difficult to reason about accurately. The second describes three types of learning that might take place when someone learns to reason about rotations: insight learning, the acquisition of intuitions, and spatial tuning. Third, we review the complete set of empirical questions that are addressed in the research to be reported (e.g., questions about individual differences). Lastly, we describe the basic experimental method that is used in the experiments (a form of interaction in a virtual environment).

1.1. Spatial organization in reasoning about rotations

Seeing that an object rotates is an example of perceptual organization. Perceptual organization may be defined as the perception of one set of relations among things when a different set of relations might have been perceived instead (see Pani, 1997, 1999). It is important to understand that neither set of relations is intrinsically correct. Rather, there is more than one way to organize the relations among things. Alternative organizations, however, may not be equally effective in furthering the successful performance of a particular task.

Examples of perceptual organization often present ambiguous pictorial displays. These tend not to demonstrate the value of perceptual organization for psychological description of the world generally. Consider some additional examples of spatial organization. When honey bees encode the location of a set of flowers, they represent it in terms of an angle relative to the sun (e.g., Schone, 1984). They do not use the visual angle between the sun and the flowers. Instead, they use an angle in the horizontal plane between the flightline to the flowers and the line along the ground between the bee and where the sun projects vertically to the Earth. When communicating this angle in the hive, the bees indicate it relative to the vertical on a wall of the hive. These are two organizations of spatial relations that are mapped together into a single elegant system. It is an effective system, but there are alternatives for encoding both locations and angles.

In human cognition, spatial organization includes using a variety of different reference systems for the determination of physical properties. For example, a cube in its typical orientation to the vertical is considered to be a "flat" shape, with parallel sides. It is perceived to be geometrically simple, and people find it easy to visualize. In contrast, when the cube is reoriented, so that a main diagonal is vertical, the properties of a cube are seen differently. The cube appears to be a "pointed" object; the "sides" of the cube are not parallel; and the overall shape is not seen to be simple at all (e.g., Holden, 1971). People have great difficulty visualizing the cube in this orientation (Hinton, 1979; Pani, Zhou, & Friend, 1997).

1.1.1. Rotation of objects considered as sets of points

A simple rotation consists of circular motion of a rigid object around a single axis fixed in space. Such a rotation creates a kinematic structure with discrete properties: there is a parallel set of circular motions of all points on the object. The motions have a single angular speed, they trace arcs that are centered around a single line in space, the axis of rotation, and the planes of these arcs are parallel to each other and perpendicular to the axis. The axis and planes of a simple rotation have an invariant orientation in the larger space that contains them. For example, the axis of rotation for a lighthouse beam is the fixed vertical axis of the tower, and the plane of rotation is the fixed horizontal plane in which the beam circles.

1.1.2. Rotation of objects considered as orientable structures

An object that rotates has a particular relationship to the axis and planes of a rotation, as illustrated in Fig. 2. To conceptualize this relationship, it is useful to think of spherical coordinates around a polar axis. For example, the orientation of a solar panel on a roof is represented in terms of slant to the vertical (sometimes called altitude) and compass direction of that slant-angle in the horizontal plane (sometimes called azimuth). Thus, the panel may be described as slanted at 45° and facing south. Gibson (1950) used the term slant for the altitude of an oriented object and called the angle within the horizontal plane direction-of-slant; we will shorten this term to slant-direction. If an axis of rotation is considered to be a polar

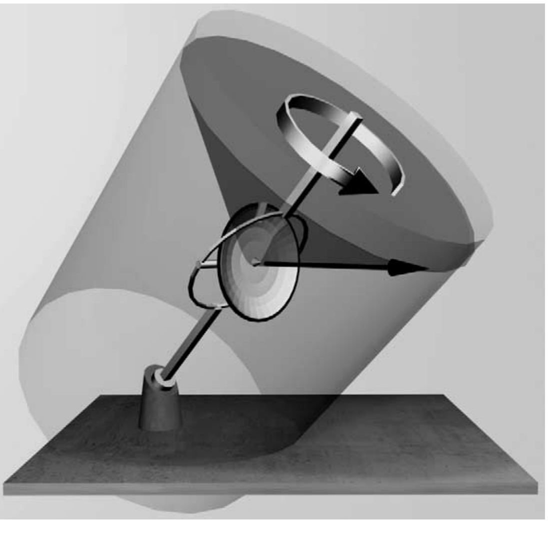

Fig. 2. An illustration of the kinematic relations in a simple rotation. The dish maintains a constant slant to the shaft, and the rotation moves this angle in a circle around the shaft.

axis, then the object that rotates will have an invariant slant to the axis, and the motion is described entirely as a continuous change of slant-direction in a circle around the axis.

When the axis and planes of circular motion that characterize a rotation are aligned neither with the intrinsic structure of the object nor with the vertical, as illustrated in Figs. 1B, 2, and 3D, the motion of the object may not be perceived to be a rotation, and it may not be represented correctly when it is posed as a problem for spatial reasoning. For example, consider a task in which an object is rotating but the axis of rotation is invisible (e.g., a rotating asteroid). If an individual perceives the object to be rotating, rather than just to be moving in some way, they will be able to indicate the orientation of the axis of rotation (Pani, William, & Shippey, 1995; see also Shiffrar & Shepard, 1991). In this task, if the rotation axis is oblique to the principal directions of the environment (e.g., the vertical) and passes through a noncanonical axis of the object, observers may view the display for nearly a minute and then indicate a direction for the axis that is incorrect by 35° or more. The motion is seen to be a continuous change of orientation, but the spatial organization of a rotation is not perceived.

The environment, with its vertical axis, is the primary reference system for the determination of the orientation of objects (e.g., Pani & Dupree, 1994; Rock, 1983). It is with respect to the vertical that one knows that a stack of boxes is well balanced or that a table is horizontal and affords the safe placement of cups of coffee. Thus, when observers look at an object that rotates around an axis that is oblique both to the object and to the environment (if the rotation is *double-oblique*), they accurately perceive its orientation relative to the vertical. However, to see the rotation, the orientation of the object must be perceived relative to the axis and planes of rotation. In other words, the vertical reference system "captures" the spatial organization of the object with regard to its orientation, and the kinematic structure of the rotation is not seen.

As noted earlier, a comparable situation results when participants are shown a static assembly (e.g., Fig. 1B) and are asked to predict the outcome of a double oblique rotation. This is a challenging problem for most people even when the axis of rotation is made visible in the form of a prominent mechanical shaft (Pani, 1993; Pani

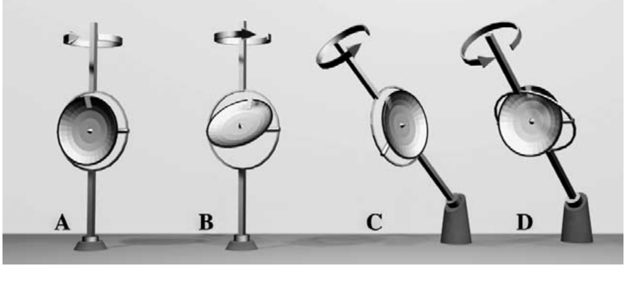

Fig. 3. Four rotations formed from varying the slant of the dish to the shaft and the slant of the shaft to the vertical.

& Dupree, 1994). The difficulty appears again to concern the orientation of the rotating object. In a rotation, the slant of the object to the shaft will be invariant, and the slant-direction around the shaft will change continuously. Despite the presence of the shaft, however, the individual is disposed to perceive the orientation of the object relative to the environment. In Fig. 1B, the dish is "upright" and "pointing horizontally, off to the right." This is a description of the orientation of the dish in the environment, and it can only stand in the way of organizing the problem effectively.

It is instructive that the other rotations in Fig. 3 are easier for people to reason about correctly. The rotation about the vertical, shown in Fig. 3B, is not challenging because the appropriate description of the dish relative to the shaft follows from the fact that the shaft is a local instance of the vertical. Note that success with this rotation also demonstrates that weaknesses in reasoning about rotation are not due to over-reliance on familiarity with common mechanical devices. People do not expect rotation necessarily to involve alignment between the objects and the rotation axes, as occurs in many familiar motions (e.g., knobs and wheels).

When there is alignment between objects and motions, as in Fig. 3C, rotations are relatively easy for people to perceive or imagine. Consider that when an object is perceived to have an orientation, there must be an intrinsic (i.e., object-relative) reference system for the object. This system defines such properties as top and front of the object, and it is this system that is perceived to be oriented to an external reference (e.g., when an object is seen to be upside down). When the dish and the shaft are aligned, the rotation axis is identical with a salient direction in the object-relative reference system. As a consequence, the slant of the object to the axis of rotation is immediately apparent in perception of the assembly.

In both the object-aligned rotation and the vertical-aligned rotation, the axis of rotation is aligned with a standard axis that is salient in perception of the scene. The slant of the object to the axis is a natural component of perception of the scene, and the individual need only consider how the changing slant-direction will move in a circle around the salient axis (also see Section 4). The rotation in Fig. 1A aligns the motion with both natural reference axes (as in a lighthouse) and is the easiest of all to perceive and to reason about.

When orientation is described in terms of a distinction between aligned and oblique, the double-oblique rotations of Fig. 3D are just one out of four classes of rotation. Consider, however, that alignment is a singular orientation. Across the set of all orientations, there are far more oblique orientations than there are aligned ones. Hence, the double-oblique rotations are quite possibly the general case (e.g., Pani, 1999; Parsons, 1995). Learning about rotational motion in general can be considered to be a broadening of understanding from the special cases favored by technological design (e.g., knobs, cranks, and wheels) to the general case of how things can move in the world.

1.2. Learning

Given the nature of rotation, it would be reasonable to expect that experience with simple rotations would lead to a development of insight about them. The term insight would mean two things. First, after an initial period of exploration, during which errors would be prevalent, individuals would reorganize their experience; they would come to understand the invariant spatiotemporal relations that were useful in predicting the outcomes of rotations (e.g., Kohler, 1929). Second, these relations would characterize rotational motion in general, and performance thereafter would be correct across the domain (Wertheimer, 1959).

Simple rotation would seem to be an excellent candidate for insightful learning. Rotation is an instance of a mathematical group: a closed system of reversible transformations. Moreover, rotation conforms to a set of elegant regularities that lend themselves to a principled formulation. Certainly kinematic principles are taught in the fields of mathematics, engineering, and the physical sciences, and experts learn to use them in their work (consider Chi, Feltovich, & Glaser, 1981; Larkin, McDermott, Simon, & Simon, 1980).

Research on naïve physics suggests that general physical principles do not necessarily reach into everyday reasoning (McCloskey, 1983; Proffitt, Kaiser, & Whelan, 1990). When they do, they may be tied to perceptual experience or to particular familiar events (Kaiser, Jonides, & Alexander, 1986). However, the question for a study of learning is what happens when naïve understanding is transcended by learning? Quite possibly, the transition from naïve to informed kinematics would include an insightful understanding of the invariant properties of the domain.

On the other hand, there is evidence in the recent cognitive literature that learning may proceed as an accumulation of episodic experiences, and the formation of relatively concrete schemas, rather than the abstraction of general principles. For example, memory of geographic layouts may favor thinking about a scene from the point of view in which it first was seen (Shelton & McNamara, 2001). The development of automaticity of attention from advanced expertise may depend on the rapid recognition of individual instances (Logan, 2002). People may reason in diagnostic situations through consideration of past cases (Brooks, Norman, & Allen, 1991); they may categorize by matching to exemplars (Medin & Ross, 1989; Nosofsky & Johansen, 2000); and they may extend analogies reliably only after a set of relations has been demonstrated across a variety of individual cases (Gick & Holyoak, 1983). In the realm of logic, problems presented as categorical syllogisms appear to be solved through the use of diagrammatic mental models (Johnson-Laird, 1983). Even reasoning that has become quite general may be couched in terms of schemas that remain closer to concrete concepts—such as the concept of permission—than to universal logical relations—such as the principle of modus tollens in conditional reasoning (Cheng & Holyoak, 1985; see also Ross, 1984). In the study of decision making, where it might be expected that universal quantitative procedures (e.g., expected utility) would be applied, researchers speak of individuals' intuitions born of concrete experience rather than a calculus over variables (e.g., Gigerenzer et al., 1999; Hogarth, 2001). Finally, the data suggesting that natural categories are represented in terms of typical instances argues that even fundamental categorical knowledge may remain closer to experience in the world than would be warranted by the extensions of the categories (e.g., Rosch & Mervis, 1975).

An accumulation of new spatial intuitions across the domain of rotations would provide an alternative to the development of insight. Intuitions, as we will speak of them, are mental representations that encode knowledge from relatively specific classes of perceptual experience. For example, an intuition might pertain to rotations around the vertical but not to rotations around other axes. Intuitions may be generated in the context of problem solving and reasoning, and their role in thinking may be expressed in logical terms. For example, an intuition might be used to disconfirm a hypothesis. However, much as a permission schema may be described in terms of the rule of modus tollens without being nearly so general, an intuition in our sense may participate in reasoning without being highly general. In addition, because it is a representation tied to perceptual experiences, an intuition may have no ready expression in the linguistic community. Such an intuition will be implicit in the sense that an individual may have little ability to verbalize the character of the representation that they are using. As a participant said in a related study, "I don't have words, I have hand gestures." (Pani, Chariker, & Fell, 2005). Learning through the development of spatial intuitions would take place as a local acquisition of useful perceptual information and the generalization of this to new instances. Where new instances were substantially different from the intuitions obtained, new perceptual information would have to be acquired. Where multiple intuitions had been developed, they might be combined into more flexible and general cognitive schemas. Such schemas typically would have neither the universal character of kinematic principles nor the conventional character of language.

A third type of learning would consist of a domain-general adaptation of spatial reasoning that would generate a gradual reduction in error homogeneously across all rotations. It might begin as a process of elimination, in which larger errors were avoided, and then become a process of spatial tuning, as answers became increasingly well-informed and accurate. Such a process would be a form of hill climbing that was implemented as a refinement of trial and error in the context of feedback (i.e., a version of operant learning). Experiments on sensorimotor adaptation (e.g., mirror drawing) appear to provide a model for such a process of learning.

1.3. Empirical questions regarding learning to reason about rotations

Our concern in this project was with learning that takes place through experience with perceived physical assemblies rather than with the mathematics taught in engineering and the physical sciences. A first question for this approach was whether learning to reason about rotations in this way would be efficient for the typical individual. The data illustrating the difficulty of reasoning for certain rotations encouraged uncertainty about this.

Assuming that learning would take place, a second question was whether it would include sudden or gradual changes in performance. Sudden changes would be clear evidence in favor of an insightful form of learning. Gradual changes might take two forms: either a steady decrease in error homogeneously across all problems or a steady increase in the proportion of problems solved correctly. A gradual and homogeneous reduction in error would imply spatial tuning. Alternatively, a gradual increase in the proportion of correct answers would imply that correct solutions were being learned in individual variants of the problems. That is, this pattern of data would support the intuition model.

These three models of learning are not inconsistent with each other. Different people could learn differently, and the same person might combine processes. Nonetheless, the different hypotheses imply distinct patterns of data in the progress of learning, and it is possible to detect these patterns when they occur.

A third general question addressed by this research concerned the information in the learning situation that would best drive or facilitate learning. First of all, we tested whether providing salient spatial information that emphasized invariant properties of rotation would accelerate learning. If participants were forming insights into the nature of rotation, this information, provided in computer graphics, should have speeded the process.

We also tested the relative value of four elementary types of learning situation. One was continued experience in reasoning without feedback or observation of the motions. Continued experience would allow participants to gauge their confidence in their answers, compare their answers across the different types of rotation, and work toward a common approach to the problems. The second learning situation included feedback and demonstration of the correct answers. This situation would be highly effective in prompting an effort to reformulate existing knowledge. The third situation permitted participants to view and control the rotations after they had reasoned about them. This situation permitted learning in which new perceptual information about the motions was added to existing knowledge. Finally, the fourth situation combined all of these sources of information. Determining the relative effectiveness of the four learning situations added substantial information about how information was being used to drive learning.

A fourth area of investigation concerned individual differences in spatial reasoning. What is the relative importance of motivational and intellectual factors in accounting for individual differences in spatial reasoning? Do the psychometric variables of IQ and spatial ability predict performance in these spatial reasoning tasks? What part of IQ would correlate with performance, and how would the magnitude of this effect compare with pure measures of spatial ability?

In part, these questions about individual differences concerned the characteristics of people who are good at spatial reasoning. Equally, however, we were interested in characterizing the nature of spatial reasoning by looking to see who was good at it. Is reasoning about rotations an effortful process in which people who work harder do better? Is reasoning about rotations a higher level process, such that people who are measured to be more effective reasoners will perform better? Is reasoning about rotation an intrinsically spatiocognitive problem, such that people who are generally good at thinking about spatial transformation will be better at it? The answers to these questions would not necessarily favor one of the models of learning over the others. However, answering these questions would shed light on how learning according to that model actually progressed.

We report the results of two longitudinal experiments conducted in virtual reality (VR) that were aimed at answering these questions. The use of VR raised an additional question that was addressed by this research. On the one hand, VR can be valuable in research, training, and education: It can provide realistic and motivating experiences relatively efficiently. On the other hand, VR is not totally realistic, and it has been our experience in discussing VR methods with colleagues that there often are serious reservations about whether learning that takes place in a virtual environment will generalize to reasoning with real objects. In the first of the two experiments, we tested whether such generalization occurs.

1.4. A framework for the study of spatial learning

To investigate the nature of learning to reason about rotations, a task was devised for Experiment 1 that was intended to provide optimal opportunities for learning under any of the three models of learning. This task included self-paced reasoning about the outcomes of possible rotations, a spatial system for indicating those outcomes, immediate feedback about whether a response was accurate enough to be considered correct, spatial demonstration of the correct answer (to avoid the need to verbalize), manipulation and observation of the rotation, and extended repetition of the task with a variety of rotations. In two variants of the task, additional spatial structures were added to the scene to encourage an understanding of the invariant properties of rotation.

The demands of such an experimental task are difficult to satisfy in a psychological laboratory unless it is implemented in virtual reality. We used what is called desktop VR, in which the perceptual experience is like looking through a window (the glass screen of the computer monitor) into a 3D spatial world. Participants interact with the world through remote control (e.g., mice, dials, and knobs). In the system that we used, the experience includes photorealistic graphics (e.g., with correct linear perspective, occlusion, shading, and specular highlights), compelling stereoviewing, smooth object motion, including the global motions that provide effective depth perception, and smooth user interaction with the scene. Objects appear realistic (e.g., formed of shiny metal), 3D spatial relations are vivid, and the environments are convincing (e.g., with a clear ground plane). People become adept at use of the controls much as one becomes adept with using rakes, screwdrivers, and handles of various kinds.

There are certain definite advantages of using VR over physical assemblies. First of all, it is possible to incorporate objects, and methods for manipulating them, that would be difficult and expensive to build. In addition, it is possible to use objects and methods that are products of thought and imagination rather than physical manufacture (so called computer visualization).

2. Experiment 1

The first experiment was designed to measure the rate of learning in reasoning about challenging rotations and to determine the nature of the learning (i.e., insight, the accumulation of intuitions, or spatial tuning). Participants observed, reasoned about, and interacted with compelling representations of concrete objects in VR scenes. Experience with the objects included repeated trials in which there were three self-paced phases: (a) indication of a possible rotation and prediction of the outcome, (b) feedback and demonstration of the correct answer, and (c) simulation of the rotation (including control and observation). There were three experimental conditions. One group of participants saw only the standard objects in the scene. Two other groups received supplemental computer visualization that was aimed at emphasizing invariant information about rotation. One group saw the computer visualization at all times. The third group did not see the visualization during reasoning; it was added to the scene during the feedback and simulation portions of the trial. At the end of the experiment, generalization of learning in VR to reasoning with real objects was tested.

2.1. Method

2.1.1. Participants

Participants were 48 university students, including 27 women and 21 men. They were paid $7.50/h for their participation. Participants were recruited by posting advertisements throughout the university campus. The advertisements explained that a study of spatial learning in virtual reality was being conducted. They indicated the rate of pay and the anticipated duration of participation. All participants had normal or corrected-to-normal visual acuity and stereovision.

2.1.2. Design and materials

2.1.2.1. Psychometric tests. Each participant completed a series of psychometric tests before the experiment began. The series included three tests of spatial ability. The Cube Comparisons test (Ekstrom, French, Harman, & Dermen, 1976) was administered because of its focus on challenging rotation problems. The Block Design subtest of the Wechsler Adult Short Intelligence scale (WASI; Wechsler, 1999) was administered as a test of visuospatial construction. The third spatial test was the Space Relations subtest of the Differential Aptitude Tests (DAT; Bennett, Seashore, & Wesman, 1989). The DAT: Space Relations is *de facto* the standard test of spatial ability (e.g., Mackintosh, 1998). It involves determining how a patterned flat surface will look after it is folded into a solid. This test is particularly relevant to the present research, as it tests the ability to reason with relatively complex spatial transformations. Each participant also completed a short version of a basic test of intelligence. This included the Matrices and the Vocabulary subtests of the WASI. The WASI Matrices is similar to the Ravens progressive matrices. It is considered a test of fluid intelligence. WASI Vocabulary is a test of crystallized intelligence. The two tests together provide an estimate of full scale IQ.

2.1.2.2. Reasoning in VR. In each trial of the experiment, participants reasoned about rotations of a transmitter dish like that seen in Figs. 1–3. By varying the height and position of the post on which the assembly stood, the center of the dish was always at the center of the scene. The dish in principle could rotate around three orthogonal axes, but participants always reasoned about rotations around the main shaft. As illustrated in Fig. 4, a ground surface, large squares arrayed on the ground, and the vertical post supporting the dish assembly provided a clear sense of an environmental reference system in a virtual world.

The displays were generated by SGI 320 computer systems specialized for the smooth presentation of interactive 3D graphics. The displays were presented stereoscopically on 21 in. high-resolution monitors (1024×768 pixels). Stereoscopic presentation was implemented with the CrystalEyesII optical system. LCD shutter glasses (which fit over corrective eyeware) alternately blocked the view of each eye. The glasses were synchronized with alternating screen images showing left- and right-eye views of the scene at 60 pairs of frames per second. This provided full screen stereoscopic vision of the scene. The integration of binocular views provided substantial antialiasing of the images. Participants were tested individually in two small rooms with the lights turned off and the doors closed.

The graphics and the capability for user interaction with the scene were generated by C++ computer programs that made calls to the Open Inventor graphics library. The participant interacted with the scene with a standard mouse that moved a cursor and with a pressure-sensitive knob, a Magellan SpaceMouse, that rested on the tabletop.

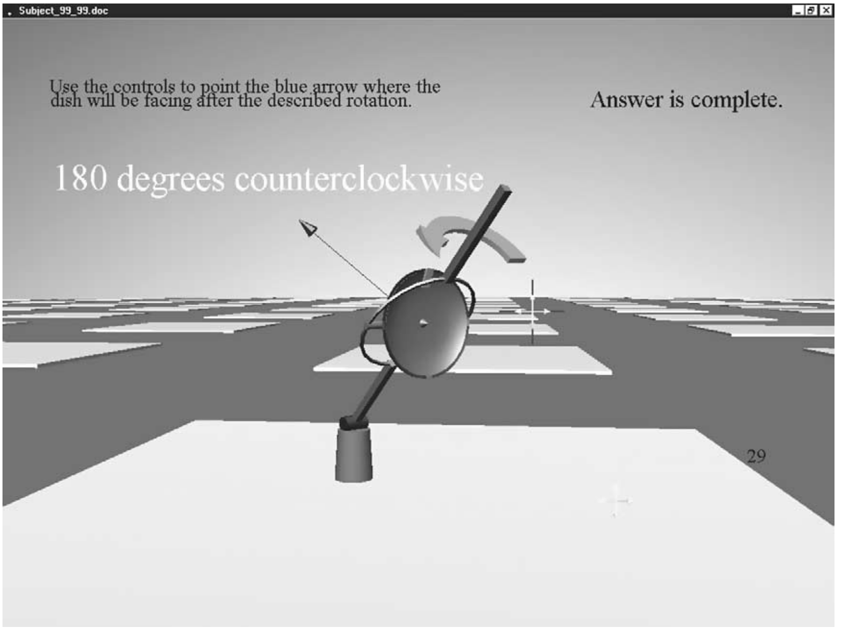

Fig. 4. Screenshot from the reasoning phase of a trial in the Standard VR or Visualization-Instruction conditions of Experiment 1. The long arrow was oriented by participants to show which direction the dish would face after the indicated rotation.

The participants were divided into three experimental conditions. In one condition, called Standard VR, participants interacted with the standard objects necessary for reasoning, response, and instruction. A second condition, called Visualization-Always, included supplemental computer visualization intended to encourage insights about invariant properties of rotation. The visualization was present throughout each trial. In the third condition, called Visualization-Instruction, the portion of a trial in which participants reasoned about a rotation was the same as in Standard VR. The additional computer visualization was present during feedback and exploration of the motion. For both of the Visualization groups, the computer visualization was withdrawn, without warning, halfway through training. That is, for the second half of the experiment, all three groups reasoned and learned in the Standard VR task.

Assignment of participants to the three groups was balanced according to three criteria of decreasing priority. The highest priority was the score on the DAT: Space Relations; the second priority was the score on overall IQ; the third was gender. Note that balancing gender across conditions would normally be a primary consideration for a test of spatial reasoning. However, such balancing is based on the fact that gender differences are found on tests such as the DAT: Space Relations. With Space Relations scores in hand, it seemed that the most effective method for balancing participants across groups would be to match participants on these scores.

2.1.2.3. Reasoning with the physical object. Generalization of learning in VR to reasoning with a real object was tested with a physical model that was built to have the same geometric properties as the virtual assembly in the VR scene. The shaft was made of brass with a 6 mm square cross-section. The ring holding the dish was a polygon identical to the corresponding part in VR. The dish was made of wood painted silver. It was 18 cm. in diameter, and the total assembly was 26 cm. long. The shaft was held in place at the bottom by wooden supports with a 9-cm. square cross-section. The supports connected to a 62-cm. wide wooden platform that sat on a tabletop aligned with the experimental room. The dish assembly was held in the same positions and orientations as those seen in the VR scene.

To indicate the outcome of a possible rotation of the physical model about its shaft, the participants held up a disc with a pointer through the middle. The disc represented the dish, and the pointer emphasized the direction it would be facing. The experimenter recorded the orientation by placing a pin in a styrofoam sphere, 13 cm. in diameter, that was suspended on a vertical rod that could be placed near the dish assembly. The sphere had latitude and longitude clearly marked in 45° intervals and was aligned with the reference frame of the rotational assembly, table, and room. The pin locations were later converted to numerical coordinates with special purpose protractors that were placed onto the sphere. This scoring system has an error in the range of $\pm 10°$ (see Pani & Dupree, 1994). Because participant errors in the task tend to be large (i.e., between 35° and 120°), this system is appropriate for statistical comparison between groups.

2.2. Procedure

2.2.1. Reasoning in VR

There were three self-paced phases in each trial of the experiment, as well as a self-paced pause between trials. The phases were problem solution, feedback and demonstration of the correct answer, and simulation of the motion. In each phase, standard instructional information was presented to the participant in the upper left-hand corner of the scene. The participant always moved to the next phase of the trial by clicking on a written statement in the upper right-hand corner of the scene (e.g., "Answer complete"). These written messages were typically presented in black letters, and they appeared to be fixed in space above and behind the objects in the 3D scene. There were two white double-arrow controllers in the scene. Each of these was composed of two crossed and joined solid arrows that could be grabbed with the mouse and moved freely in the frontal plane of the scene. The lower of the two controllers could be used to change the participant's viewpoint within a moderate range (a total of 30°) both horizontally and vertically. This allowed disocclusion of object parts and provided motion parallax that reinforced the visual realism of the virtual world.

In the first phase of a trial, problem solution, the dish assembly was presented in a static orientation and a written instruction said to "Use the controls to point the blue arrow where the dish will be facing after the described rotation." A frame from this phase of a trial is presented in Fig. 4. A particular rotation was indicated with a curved arrow located around the shaft and a bright yellow written message (e.g., "90° clockwise"). Because earlier tests revealed that participants sometimes neglected this message, it was programmed to jump in size 1 s after the trial began. The participant indicated the direction the dish would face after the rotation by pointing a blue arrow in that direction. The arrow always extended out from the center of the dish. To point the arrow, the participant grabbed the upper double-arrow controller with the mouse and moved it in the plane of the screen. As the controller was moved up and down, the direction that the arrow pointed rotated up and down relative to the vertical. As the controller was moved left and right, the direction that the arrow pointed swung around the vertical.

When the participant was satisfied with the direction that the blue arrow pointed, they clicked on the words "Answer complete." in the upper right-hand corner of the scene. That initiated the second phase of the trial. Written feedback appeared on the screen in bright white letters, saying either, "Your answer is correct (or at least close)." or "Your answer is not close enough." At the same time, a red arrow appeared in the scene indicating the correct answer (i.e., where the blue arrow should have been pointed). A written message under the feedback said, "The red arrow shows the correct answer." The participant thus had an opportunity to compare their own answer with the correct answer in the context of the feedback. Positive feedback was given when the participant was within 30° of the correct answer (see Pani, 1993; Pani & Dupree, 1994).

When the participant was finished reviewing the feedback, he or she clicked "Done with feedback." At that point, the participant twisted the SpaceMouse to generate a simulation of the rotation in question. The virtual object rotated smoothly in either direction for as long as the participant kept pressure on the dial, with greater pressure producing a faster rotation. The red arrow indicating the correct answer to the problem (but not the blue response arrow) remained in the scene. While the dish pointed in the direction of the correct answer, the red arrow turned bright yellow. The trial could not be advanced to the next phase until this occurred at least once. When the participant finished viewing the rotation, he or she clicked on the message, "Done with dial." The program then paused, with an appropriate message on the screen. The participant was free to take a break at that point, and then initiated the next trial by clicking on the message "Next trial."

The procedure just described was used in all three conditions. The computer visualization that was added to two of the conditions included a translucent green surface known as a surface of revolution. This surface surrounded the main shaft of the assembly and illustrated the set of directions that the dish would face throughout the indicated rotation, as illustrated in Fig. 5. Such a surface demonstrates that the slant of the dish to the shaft will be invariant throughout the motion and implies that the motion will take place as a circular movement around the shaft. This surface was demarcated with black lines that indicated the outcomes of all rotations that could

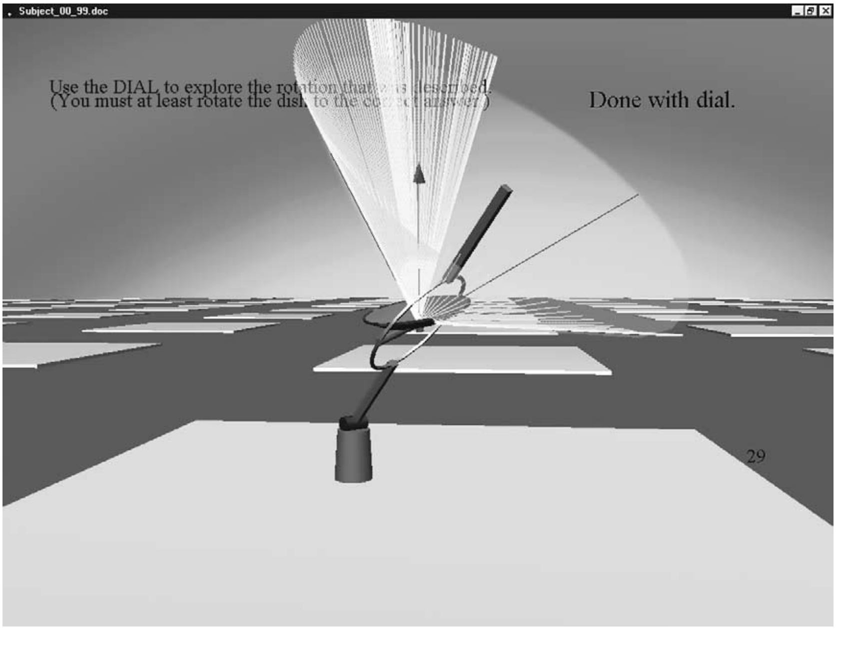

Fig. 5. Screenshot from the simulation phase of a trial in the Visualization-Instruction or Visualization-Always conditions. The cone was a translucent green surface of revolution that showed everywhere that the dish would face throughout the indicated rotation. The cone was demarcated in 90° intervals. As the participant moved the object through the rotation, the lines (bright yellow) traced the directions that the dish had faced.

take place in 90° intervals. During the simulation phase of a trial, when the participant controlled the motion of the assembly, a series of bright yellow lines left a trail of every direction that the dish had faced during the rotation. As the dish rotated, this trail progressively traced out the green surface of revolution.

The Visualization-Always group saw the green surface of revolution during every phase of a trial. That is, it was available during reasoning about the rotation as well as during feedback and simulation. The Visualization-Instruction group reasoned with a scene that was the same as in Standard VR. The surface of revolution appeared with the presentation of feedback and continued through the simulation of the motion. As noted earlier, for both of these groups, the computer visualization was withdrawn, without warning, halfway through training.

Data recorded by the computer in each trial included the time spent viewing the problem before the blue response arrow was moved, the time elapsed for each phase of a trial, and the total elapsed time (not including pauses). All positions of the response arrow were recorded as well as all positions of the mechanical assembly during simulation of the motion.

Although the primary focus of this experiment was to determine the nature of learning for the more challenging rotations, a systematic variety of rotation problems was presented. This was done because there is much information about how to think about the challenging rotations from considering the easier ones. In addition, the easier problems provided a natural control condition for assessing the generalized improvement in reasoning about rotations of all types and the increasing mastery of the task as a whole.

The rotation problems were organized into blocks of 32 trials. The 32 trials were generated by a factorial combination of the orientation of the shaft in the environment, the orientation of the dish to the shaft, and the amount of rotation. The orientation of the shaft in the environment was vertical, frontal–horizontal, partially oblique (the rotation axis was in a Cartesian plane of the environment but oblique at 45° to the Cartesian axes in that plane), or fully oblique (the rotation axis was oblique at 45° to all Cartesian axes and planes of the environment). The orientation of the dish to the shaft was perpendicular, oblique at 45° "up" the shaft, parallel to the shaft (i.e., facing straight out), or oblique at 45° "down" the shaft. The amount of rotation was either 90° or 180°. Finally, so that individual rotation problems would be repeated as seldom as possible, additional types of display variables were included. In particular, there were six partially oblique axes to choose from, four fully oblique axes to choose from, a particular type of rotation could begin at any one of eight starting orientations around the shaft in 45° intervals, and the direction of rotation could be either clockwise or counterclockwise. A computer program generated the trial sequences so that no rotation problem was repeated before it was necessary. The first repetition of any problem occurred on trial 513 (halfway through the experiment), and the first repetition of a double-oblique rotation occurred on trial 772 (more than three quarters of the way through the experiment). Only 27 of the double-oblique rotations repeated in the entire experiment, and the repetitions were spread out evenly between trial 772 and the end of the experiment (trial 1024). No problem was repeated twice.

The instructions for the VR learning task were in three parts. First, the participants were shown rotations of the dish around the shaft for a variety of orientations of the dish to the shaft and the shaft in the environment. The variations of orientation were emphasized, and it was pointed out to the participant that their task would be to predict the outcomes of rotations such as they were being shown. Second, the participants practiced orienting the blue arrow. In 32 separate trials, a red arrow systematically sampled all directions in 3D space. The participant used the mouse and white controller to match the orientation of the blue arrow to the red arrow. When the orientation matched, the red arrow turned yellow and the participant was able to move to the next trial (with a mouse click). Finally, the participant was given practice with the experimental task that they would be performing. There were three practice trials: one with a vertical shaft and an oblique dish, one with an oblique shaft and a parallel dish, and one with an oblique shaft and an oblique dish. The experimenter then invited further questions and verified that the task was well understood.

Participants solved rotation problems for an hour per day until they had completed the total 1024 trials (32 blocks of trials with 32 trials in each block). This generally required about 8 days of testing spread over 2–3 weeks. At the completion of the learning in VR, participants were tested with a series of rotation problems with the physical model.

2.2.2. Reasoning with the physical object

At the completion of the VR procedure, an additional day of testing took place with the physical assembly. A control group, which had not participated in the VR training, completed the identical procedure during the same time period in which the VR groups were being tested. Instructions for the VR-trained participants were that the task with the physical object was very similar to what they had been doing in VR. They were told to use what they had learned in VR to solve this new set of problems. They practiced the new experimental task with two displays before proceeding to the experiment. Instructions for the control group were similar to the instructions that the VR-trained participants received at the beginning of their training and to instructions used in prior research (Pani, 1993; Pani & Dupree, 1994; Pani et al., 1995). With use of the physical assembly, it was demonstrated systematically that the dish could be at different orientations to the shaft and that the shaft could be at different orientations in the environment. Seven rotations with different configurations of these orientations were carefully demonstrated. It was explained that the participant's task would be to view a static assembly and to indicate the direction the dish would face after a rotation. As with the VR-trained groups, the participant then practiced the experimental task with two displays before proceeding to the experiment. The experimenter verified that the task was well understood and answered any questions the participant had.

The experimental procedure was essentially a replication of previous studies (e.g., Pani, 1993; Pani & Dupree, 1994). Rotation problems were presented one at a time, and participants took as long as they wished to indicate an answer. When they were ready, they held up the disc and pointer to indicate how the dish would be oriented after the rotation. The experimenter recorded the response time with a stopwatch and coded the answer on the styrofoam sphere. Participants reasoned about the outcome of 16 different rotations. The set of rotations was composed from a factorial combination of four orientations of the dish to the shaft (parallel, perpendicular, oblique up, and oblique down) and four orientations of the shaft in the environment (vertical, frontal–horizontal, partially oblique, and fully oblique). For each orientation of the shaft, rotations were either 90° or 180°. There were two sets of problems used across the sets of participants. The assignment of an amount of rotation (90° or 180°) to individual rotation problems was counterbalanced across the two sets. A single trial order was adopted, with the early trials containing more of the rotations known to be easier and the later trials containing more of the rotations known to be difficult.

2.3. Results and discussion

2.3.1. Learning in virtual reality—basic results

Error in each trial was calculated as the shortest angle between the participant's setting of the blue response arrow and the correct answer. The primary measure of response time was total time per trial (not including pauses between trials). Mean error and mean response time are presented in Fig. 6, broken down by the three experimental conditions, the four classes of rotation, and successive trial blocks. Clearly, the pattern of results from the Visualization-Always condition was qualitatively unique. Performance in the Standard VR and Visualization-Instruction conditions did not differ for any of the four types of rotation ($F < 1.0$). Data from these two conditions were combined in further statistical analyses. We will begin by describing the results for these two conditions and then return to discussing Visualization-Always and the effects of providing supplemental computer visualization.

Please note that for the rotations in which the dish initially faced straight up the shaft, the dish would face straight up the shaft throughout the rotation. These problems soon became trivially easy, and the data from those rotations are not included in the graphs or analyses that are presented. Comparable rotations that are included are those with the dish facing straight out from the shaft.

2.3.1.1. Type of rotation. Performance in the early trials of the Standard VR and Visualization-Instruction conditions replicated the results of previous studies using physical objects for both error and response time. Thus, performance was better when the dish was aligned with the shaft rather than oblique to it (error: $F(1,23) = 59.817$, $p < .001$; time: $F(1,23) = 96.081$, $p < .001$) and when the shaft was vertical rather than oblique in the environment (error: $F(1,23) = 50.675$, $p < .001$; time: $F(1,23) = 128.523$, $p < .001$). Performance when the shaft was vertical was superior to when it was frontal–horizontal, for the cases in which the dish was oblique to the shaft (error: $F(1,23) = 20.7$, $p < .001$; time: $F(1,23) = 56.4$, $p < .001$). Relative to the other types of rotation, participants took a very long time and committed a high level of error for the double-oblique rotations; this was supported statistically by an interaction between the two orientation variables (error:

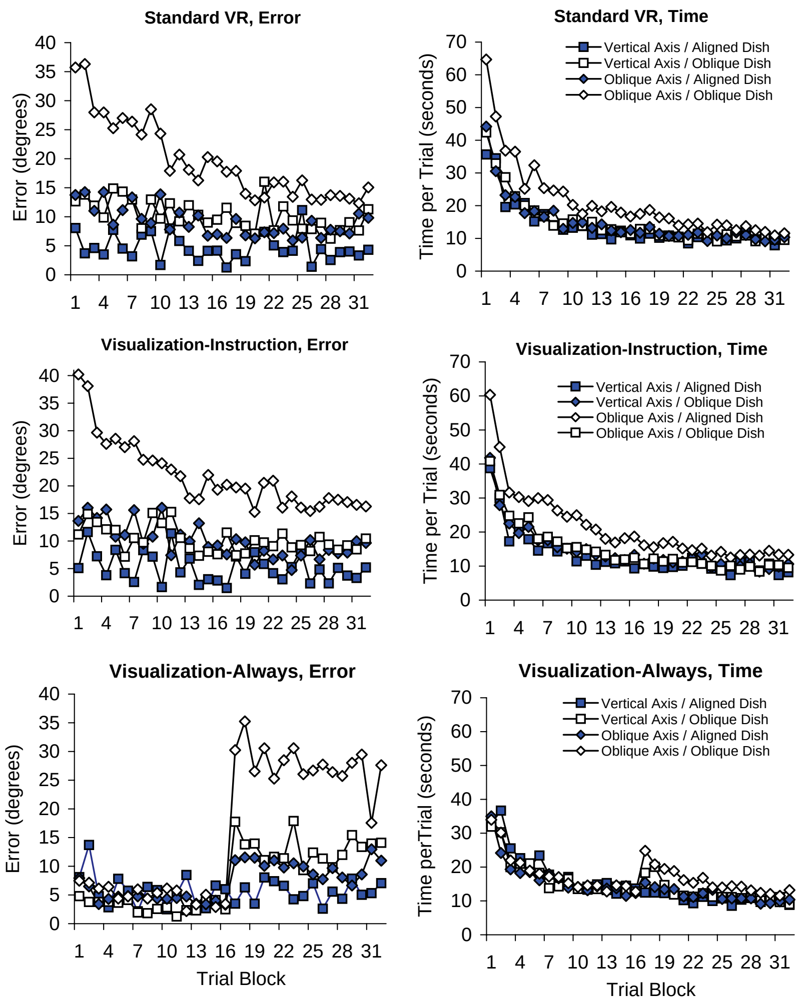

Fig. 6. Mean error and time per trial in Experiment 1, broken down by the type of rotation, trial block, and the three experimental conditions.

$F(1,23) = 17.7$, $p < .001$; time: $F(1,23) = 62.148$, $p < .001$). (All analyses of variance use the Huynh–Feldt correction for violation of sphericity in small samples.)

2.3.1.2. Learning. At the outset of this research, there was a question as to whether learning to reason well across the entire domain of simple rotations would be reasonably efficient. Experiment 1 demonstrated that for the most part it is. Substantial learning took place across all types of rotation, as indicated by a main effect of trial block (error: $F(31,713) = 16.663$, $p < .001$; time: $F(31,713) = 89.3$, $p < .001$). More importantly, there was near convergence across trials in the levels of error and response time for the different types of rotation. Thus, there was a three-way interaction of the two orientation variables (that defined the types of rotation) with trial block (error: $F(31,713) = 5.207$, $p < .001$; time: $F(31,713) = 4.7$, $p < .001$). The learning curve for the more challenging (double-oblique) rotations conformed to a classic power function with negative exponent (error in degrees as a function of trial block $= 0.44\ \mathrm{x}^{-0.31}$).

2.3.2. Three models of spatial learning

The group means presented in Fig. 6 indicate that learning took place, but they say little about the nature of that learning. For example, Fig. 6 appears to demonstrate gradual learning, but if different participants gained insights during different trial blocks, the group means might take that form. A graphical examination of individual participant data, in fact, strongly suggested that the intuition model of learning was correct. Performance of individual participants consistently switched between correct performance and large errors, with the proportion of correct responses increasing over time. This variability is illustrated in Fig. 7, which presents individual trial data for error on double-oblique rotations for three of the participants who showed substantial amounts of learning.

2.3.2.1. Patterns of change across trials. A direct quantitative comparison of the three learning models focused on predictions of the models for how error should change across blocks of trials. The change in error from one trial block to the next is given by the first derivative of error. Note that within each block of trials there were eight double-oblique rotations, and these eight rotations comprised subtypes of rotation that were used consistently. For example, a 90° rotation with the dish facing upward on a fully oblique shaft was one subtype of rotation that was presented in every trial block. Calculations of change in error across trial blocks were carried out individually on the eight subtypes of rotation. This should have maximized the similarity of the rotations across trial blocks and provided the most stringent test of the intuition model. It is important to realize, however, that there was substantial physical variation among the problems within each of the subtypes of rotation. For example, the particular orientation of the shaft could differ within a subtype of rotation (e.g., tip forward rather than backward), and the starting orientation of the assembly could take eight different angles (e.g., face forward or to the side).

The overall change in error from one trial block to the next reflects the average of changes that occur at the level of individual trials—the eight rotation problems in each block of trials. The three learning models make quite different predictions about how the overall derivative is related to the derivatives for each of the subtypes of rotation. In particular, the derivatives for the subtypes of rotation can be divided into those that are positive in a particular trial block, reflecting an improvement in performance, and those that are negative—reflecting a decrement. The insight and spatial tuning models share the assumption that the process of learning is reflected entirely in positive derivatives—those that reflect improvement. Any negative derivative, reflecting a decrement in performance, is due only to influences on performance that can be considered noise. The intuition model, on the other hand, predicts

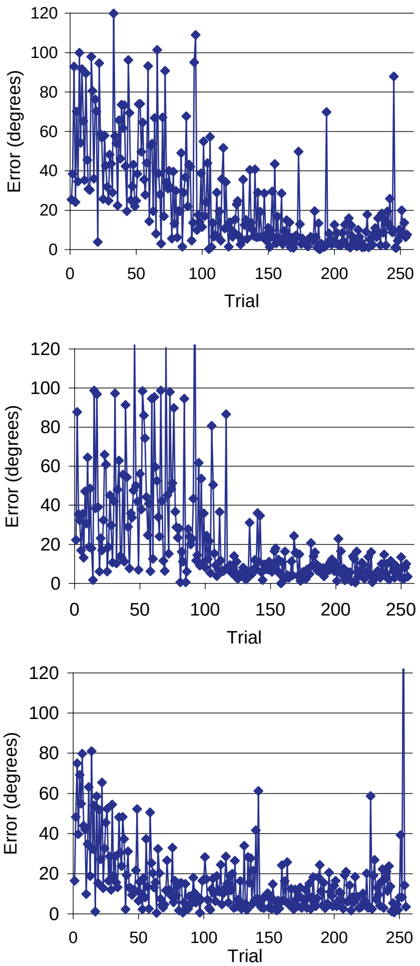

Fig. 7. Error viewed trial by trial for three representative successful learners from the Standard VR or Visualization-Instruction conditions.

a decrement in those cases that a correct answer is followed by a new problem that is not perceived to be familiar. This will be relatively common and will produce negative derivatives that are large. Thus, to account for a given overall derivative across trial blocks, the intuition model generates a relatively wide separation between the positive and negative derivatives in comparison to the predictions of the other two models.

Fig. 8 illustrates the predictions of the three models of learning for changes in the level of error. The left hand panels of the figure show the predicted derivatives for ideal individuals trial by trial. The right hand panels are different in two ways. First, they reflect data that is averaged across trials within a block and across participants.

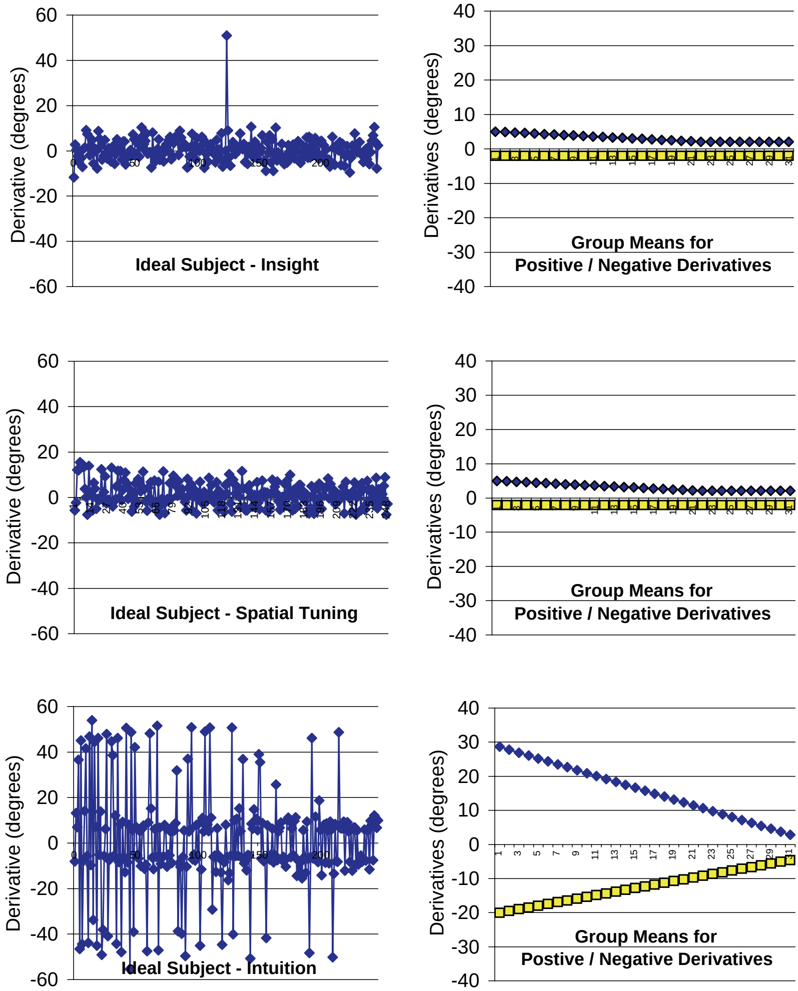

Fig. 8. Predictions of the three learning models for the positive and negative derivatives of error on double-oblique rotations: at the left, derivatives for ideal individuals across individual trials; at the right, derivatives averaged over participants and rotations within a trial block.

Second, these panels graph separately the mean derivative for those subtypes of rotation for which performance improved in any trial block, and the mean derivative for those subtypes of rotation for which performance deteriorated in any trial block. It is assumed that there are small levels of noise in the data.

To describe the data in terms of the change in error across trial blocks, each participant's raw score error from the double-oblique rotations was transformed to the first derivative of error across trial blocks. This was done separately for each of the eight subtypes of rotation within the set of double-oblique rotations. Within each trial block, the positive derivatives and the negative derivatives were averaged separately.

Fig. 9 presents the mean derivatives across participants for the positive, negative, and overall changes in performance across trial blocks. This breakdown of the change in error reveals that the small gradual improvement in mean performance across trials masked a substantial tension between large improvements in performance and large deteriorations of performance within each block of trials. Indeed, over the experiment as a whole, the mean positive change in error was 9.8 times as large as the overall change in error in the same block of trials.

This pattern of data is consistent with the intuition model, in which learning is characterized by an increase with experience in the proportion of correct responses. However, both the insight and the spatial tuning models might be presumed to be reasonable statistical fits to the data as long as suitable levels of noise were included in the models. To test this possibility, mathematical simulations of error performance, with noise parameters adjusted to the actual data, were constructed in accordance with each model.

All three of the simulations included a noise parameter. This was simulated in each trial as a random selection of an answer to the rotation problem from a homogeneous distribution of orientations in 3D space that was centered on the correct answer. For all three of the learning models, the initial range of this distribution of responses (i.e., the maximum angular distance from the correct answer) was adjusted so that the mean error for these randomly selected responses would be equal to the mean error of the actual participants in trial block 1. Of course, it was this random selection from a distribution of responses that provided the insight and spatial tuning models an opportunity to simulate the high variability in individual participant data.

The insight simulation included a step function for each participant that moved from a high level of mean error to a low level of mean error at a single trial block

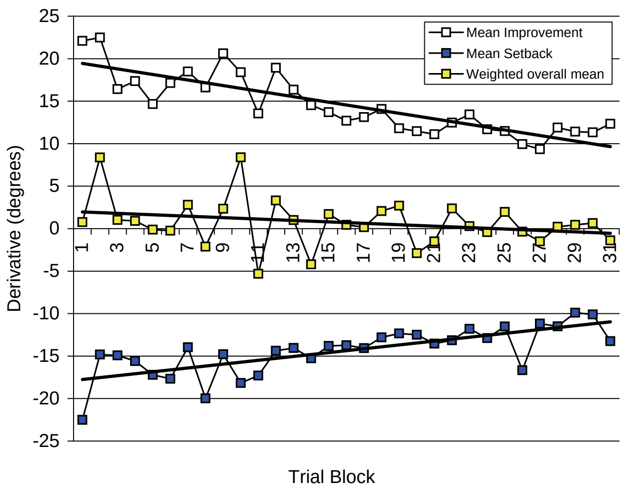

Fig. 9. Derivatives of error from the Standard VR and Visualization-Instruction conditions. The derivatives are averaged over participants and rotations within a trial block, separately for positive and negative derivatives (i.e., for improvements and setbacks).

at some time during the experiment. That is, the range and variance of the random distribution of responses around the correct answer shrank substantially after the simulated insight. The range of the smaller distribution of orientations was set so that the mean error for responses from that distribution would be equal to the mean error of the actual participants at the end of the study. To match the power function for the group means, different simulated participants achieved insight at different trials.

The spatial tuning model was simulated as a gradually decreasing range and variance of the distributions of responses around the correct answer. The range of the distribution was set so that a random selection from the distribution would generate the mean error observed in the actual data.

The intuition model was somewhat similar to the insight model, in that responses could be sampled randomly from two homogeneous distributions of orientations with either a large or a small range and variance. The same distributions of orientations were used for the intuition model as were used for the insight model. In the intuition model, however, whether a response was generated from the high-variance or low-variance distribution was determined probabilistically from trial to trial. The probability of a response coming from one or the other distribution was related to the power function that described mean performance in the actual data.

Twelve simulated participants with different random selections of responses were generated for each learning model. Each simulated participant was given the same number of problems and the same number of trial blocks as the actual participants had been given. The simulated data were converted to the first derivative of error and partitioned into positive and negative changes in performance. A mean positive change and negative change were calculated for each simulated participant in each trial block. These values are presented in Fig. 10 for the three simulations and for the actual data.

To compare the simulated and actual patterns of learning statistically, the mean positive and negative derivatives were converted to absolute values. The sets of absolute values reflected the degree to which both positive and negative changes in performance were extreme. A Tukey HSD was conducted on these values from the three simulated data sets and the actual data. The actual data were clearly different from the simulated data derived from the insight and spatial tuning models ($p < .001$). The actual data were very similar to the simulated data from the intuition model ($p = 1.000$). The data from the insight and spatial tuning models were not different ($p = .974$). Once again, this is clear support for the intuition model.

2.3.2.2. Item effects. The applicability of the intuition model implied the possibility of item effects. That is, there might be identifiable variations in difficulty among the problems that would contribute to instability in generalizing among them. To detect item effects, the displays were described along a number of physical parameters, and multiple regression was used to determine whether these parameters accounted for differences in performance. One parameter that has been shown to be influential in prior research is whether or not the dish is aligned with a Cartesian axis of the environment at the beginning of a double-oblique rotation (Pani & Dupree, 1994). The

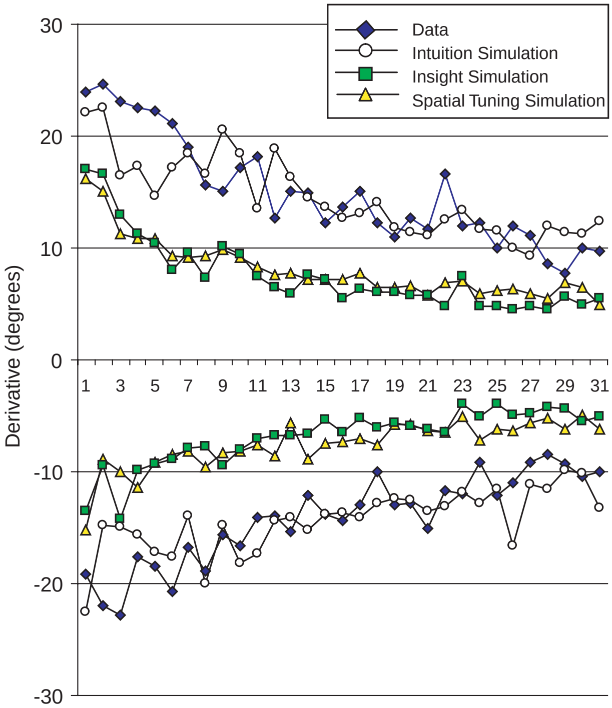

Fig. 10. Derivatives of error from computer simulations of the three learning models and the data from the Standard VR and Visualization-Instruction conditions. The derivatives are averaged over participants and rotations within a trial block, separately for positive and negative derivatives.

rotations are more difficult when the dish is Cartesian in the environment, because this orientation makes the environmental reference system more salient (when orientation to the axis of rotation is what matters). Additional parameters that were included as potential predictors in the regression analysis were whether the dish faced up or down the shaft, whether the dish initially faced frontward, sideways, or to the back of the scene, the amount of rotation, the direction of rotation, and whether the shaft was partially or fully oblique in the environment. (Some of these parameters contributed to defining the 8 subtypes of rotation and some did not.) For completeness, the trial block in which the rotation problem was presented also was entered as a predictor in the analysis.

Stepwise multiple linear regression was conducted for error and response time averaged across participants for the 256 challenging rotation problems presented in the experiment. Results of the regression analysis for error indicated four item effects (in addition to the effect of trial block). In the order in which variables entered the stepwise regression, item effects included the amount of rotation (with 180° more difficult than 90°; $r = .376$), whether or not the dish began the rotation in a Cartesian orientation with respect to the environment ($r = .178$), whether the dish faced up or down the shaft (with up being easier; $r = .124$), and whether the dish was partially or fully oblique in the environment (with partially oblique being easier; $r = .075$), $F(5, 255) = 69.9$, $p < .001$. For response time, there was one item effect: whether or not the dish began the rotation in a Cartesian orientation with respect to the environment ($r = .170$), $F(2, 255) = 204.3$, $p < .001$. The presence of item effects clearly supported the intuition model.

2.3.2.3. Individual differences and the learning models. There were substantial individual differences in this experiment (see later discussion). For example, 5 out of 24 participants had mean error for the challenging rotations below 25° in both of the first two blocks of trials. It seemed possible that those participants formed domain-general descriptions of constraints on rotation consistent with the insight model. To investigate this possibility, the raw error data from those participants were examined trial by trial for the challenging rotations in the early trial blocks. If insightful learning had occurred, there should have been instances in which a low level of error (e.g., 15° or 25°), once reached, indicated that no further large errors (e.g., greater than 45°) would be committed for a reasonable period of time (e.g., 10 trials). This pattern did not occur for any participant.

2.3.2.4. Discussion of the learning models. Learning to reason about rotations was gradual, even at the level of individual participants, and the pattern of learning fit what we have called the intuition model. In this model, the individual discovers useful information in particular cases and then generalizes this success to additional instances that are encountered in the domain. When new instances are sufficiently dissimilar to the old ones that they are not handled well, new information must be encoded for them as well. The process continues until the domain has been covered entirely in terms of a set of useful intuitions, or until they are combined into a schema that can handle a broader set of cases.

The core evidence for this model was of four kinds. First, individual subject data were highly variable from trial to trial for comparable types of rotation. Second, the course of progress across trial blocks was a very good fit to simulated data derived from the intuition model. In the simulation, correct performance often alternated with large errors, with the probability of a correct answer increasing with practice. In contrast, the experimental data did not fit simulations of an insight or a spatial tuning model, even when substantial noise was added to them. Third, there were numerous item effects within the set of challenging rotations. Thus, for example, 180° rotations were more difficult to learn than 90° rotations; problems in which the dish began in alignment with the environment were more difficult than problems in which the dish was oblique. Finally, the one participant comment that became commonplace in this experiment was that the problems were solved by remembering earlier ones. This was true even when individual problems had not been repeated.

2.3.3. Computer visualization

2.3.3.1. Results. The inclusion of supplemental computer visualization in two of the experimental conditions was intended to facilitate movement toward any possible insight during learning. The results of the experiment suggest that for the Visualization-Instruction condition, computer visualization was irrelevant to learning. In the Visualization-Always condition, computer visualization was a relative disadvantage. During early trials, when the visualization could be used to generate answers, performance was nearly perfect. When the visualization was removed in later trials, participants had learned almost nothing (for comparison of the two sets of trial blocks, error: $F(1,11) = 30.43$, $p < .001$; time: $F(1,11) = 52.58$, $p < .001$).

2.3.3.2. Discussion. It was necessary in the Visualization-Always task for participants to attend to the assembly and to think about a rotation. It was at least necessary to consider the starting orientation of the assembly, whether the motion was 90° or 180°, and the direction of the rotation. In that context, participants manually produced correct answers and then viewed the actual motion for 128 separate trials. After the computer visualization had been removed, it would seem to have been a straightforward strategy to recall what had been done when the visualization was present. The very poor performance suggests that participants in this condition had not paid attention to the important relations between spatial properties, such as the angle between the dish and the shaft, that were associated with their earlier answers.

Disconfirmation of an insight model was further supported by the ineffectiveness of adding to the learning situation computer visualization that illustrated the invariant properties of rotational motion. By all accounts, the visualization was intuitive and compelling. However, it did not accelerate learning. This null effect in conjunction with the absence of domain-general learning suggests that an abstract geometric understanding of the rotations was not at the heart of learning for most participants. This is especially interesting in light of the clear applicability of such a geometric formulation in the domain of rotational motion.

Performance in the Visualization-Always condition demonstrated a cognitive phenomenon related to change blindness (see, e.g., Simons, 2000; also Pani, 2000, for a discussion of change blindness relevant to the current data). In change-blindness, visual properties that are seen but not attended are not remembered well enough for the individual to notice changes to those properties a moment later. In the Visualization-Always condition, 128 trials of correct performance with the challenging rotations failed to establish knowledge or memory that would be useful for problem solution when the explicit guidance for the responses was removed.

2.3.4. Individual differences

2.3.4.1. Results. Individual differences in the ability to reason about rotations were large. Indeed, the range of performance due to differences among the participants was comparable to the range due to the type of rotation, as illustrated in Fig. 11. This figure shows the breakdown by participant of mean error for the challenging rotations in the Standard VR and Visualization-Instruction conditions. Despite the substantial learning that took place, some participants ended the experiment with higher mean error than other participants began it. For particularly skilled individuals, it was necessary to look inside the trials of block 1 to find high levels of error.

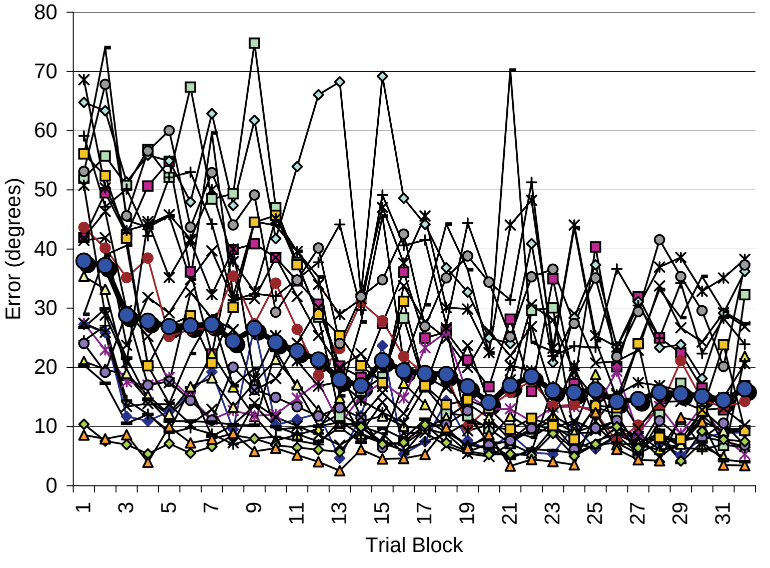

Fig. 11. Mean error for the double-oblique rotations in the Standard VR and Visualization-Instruction conditions, broken down by trial block and individual participant. Group means are indicated by the large circles.

Even then, the difficulty of the rotations for certain individuals appeared primarily in the form of large initial response times (e.g., 120 s).

The need in this task for reasoning and effortful learning suggested that the level of effort participants were investing in the task might play a role in their levels of performance. Therefore, a measure of mastery motivation was developed to complement the psychometric measures that had been collected. Because all aspects of the trials were self-paced, response time should have been related to motivation. However, response time alone was insufficient for assessing motivation, because different individuals had differing levels of need to spend time on the problems. An appropriate measure of motivation and effort was available in the relation between response time and the level of error for each individual. That is, an individual's error was communicated immediately through feedback and demonstration of the correct answers. More motivated participants could increase their time spent on the problems to drive this error value down. A measure of Mastery Motivation was calculated by taking each participant's mean response time, dividing it by the mean error, and then rescaling all of the scores so that the group mean was equal to 1.0. A high value of Mastery Motivation (e.g., 2.0) indicated a high degree of motivation to perform well.

Correlation matrices relating the psychometric measures and Mastery Motivation to mean error on the challenging rotations are presented in Table 1 for the Standard VR and Visualization-Instruction conditions. Every one of the raw correlations at least approached statistical significance ($p < .1$). Significant correlations spanned

Table 1
Correlation matrix (Pearson) relating error on the double-oblique rotations to the psychometric measures and mastery motivation for the Standard VR and Visualization-Instruction conditions in Experiment 1

	Vocab	Matrices	IQ-percent	DAT: Space	Block	Cubes	MasMot
Error	−0.327	*−0.438	*−0.388	** −0.525	*−0.338	−0.278	***−0.712
Vocab		*0.403	***0.793	*0.444	**0.491	**0.466	0.325
Matrices			***0.826	0.221	*0.375	**0.523	*0.389
IQ-percent				0.326	**0.499	**0.489	0.326
DAT:Space					***0.709	**0.552	*0.351
Block						***0.610	0.198
Cubes							0.140

Individual entries correspond to mean error, WASI: Vocabulary, WASI: Matrices, WASI-IQ percentile, DAT: Space Relations, WASI Block Design, Cube Comparisons, and Mastery Motivation. Asterisks indicate statistical significance (uncorrected) with * for $p < .05$, ** for $p < .01$, and *** for $p < .001$. $N = 24$.

the types of measures, including general intelligence (WASI: Matrices: $r = -.438$, $p = .016$), spatial ability (DAT: Space Relations: $r = -.525$, $p = .004$), and Mastery Motivation ($r = -.712$, $p < .001$). The single highest correlation was for Mastery Motivation, which accounted for nearly 50% of the variance among the participants' mean error. The predictors were quite often correlated with each other.

Organizing a large set of intercorrelated predictors into a single description requires a conceptual system to guide multiple regression (e.g., Cohen & Cohen, 1975). Intuitively, it seemed that there were three types of measures: those related to general intelligence, those related to spatial ability, and mastery motivation. A factor analysis on all 48 participants (including those who only reasoned with the physical model) confirmed this view. A two dimensional solution captured 70% of the variance among the predictors. The three tests of visuospatial cognition clustered closely together. Relatively far from this cluster was a second grouping of IQ and Matrices. The Verbal test fell in between the two clusters. Mastery Motivation moved off to the side of this Spatial-IQ axis and was closer to IQ. It was decided then to enter the measures into a hierarchical multiple regression that accounted for mean error in three blocks of predictors. The first block included Matrices, Vocabulary, and IQ. The second block included Space Relations, Cube Comparisons, and Block Design. The third was Mastery Motivation. Stepwise regression was used within each block. The ordering of blocks constituted a movement from general to specific. How much variance in mean error could be accounted for by general measures of intelligence? How much additional variance was accounted for by the more specialized skills of spatial ability? How much further variance was accounted for by participants' effort measured in this particular experiment?

In the first block of the multiple regression, Matrices emerged from the stepwise regression. This measure accounted for 19.2% of the variance among the participants on mean error, $R = .438$, $F(1,22) = 5.216$, $p = .032$. In the second block of the regression, Space Relations emerged from the stepwise calculation. This measure accounted for an additional 19.3% of the variance, $R = 0.620$, $F(1,21) = 6.585$, $p = .018$. Together these two psychometric variables accounted for 38.5% of the

Table 2
Correlation matrix (Pearson) relating error on the double-oblique rotations to the psychometric measures and mastery motivation for the Visualization-Always condition in Experiment 1

	Vocab	Matrices	IQ-percent	DAT: Space	Block	Cubes	MasMot
Error	**−0.771	−0.353	***−0.808	**−0.744	**−0.730	−0.349	*−0.563
Vocab		0.195	***0.877	**0.728	** 0.662	*0.580	0.317
Matrices			*0.624	0.040	0.035	−0.030	0.040
IQ-percent				*0.599	*0.539	0.431	0.264
DAT:Space					***0.856	*0.554	0.402
Block						**0.639	0.348
Cubes							0.253

Individual entries correspond to mean error, WASI: Vocabulary, WASI: Matrices, WASI-IQ percentile, DAT: Space Relations, WASI Block Design, Cube Comparisons, and Mastery Motivation. Asterisks indicate statistical significance (uncorrected) with * for $p < .05$, ** for $p < .01$, and *** for $p < .001$. $N = 12$.

variance among the participants on mean error. In the third block of the multiple regression, Mastery Motivation accounted for an additional 23% of the variance among the means, $R = .784$, $F(1,20) = 11.93$, $p = .003$. Altogether, these three variables accounted for 61.5% of the variance among the participants on mean error, $F(3,23) = 10.63$, $p < .001$. Similar analyses conducted just on the early or the late trial blocks generated very similar results.

A somewhat different picture emerged from relating performance to measures of individual differences for the Visualization-Always condition, as indicated in Table 2. In the raw correlations, the relation of error to the test of Vocabulary ($r = -.771$, $p = .002$) and to IQ were quite high ($r = -.808$, $p = .001$). In the same hierarchical multiple regression used for the other conditions, IQ (rather than Matrices) entered the analysis in the first block, accounting for 65.2% of the variance among the error means, $R = .808$, $F(1,10) = 18.77$, $p = .001$. No spatial measure entered the analysis in the second block. Mastery Motivation entered the regression in the third block, accounting for an additional 13.1 percent of the variance, $R = .885$, $p = .044$. The two variables together accounted for 78.4% of the variance among the means. In comparing these results to those from the Standard VR and Visualization-Instruction conditions, it should be considered that this analysis used half as many participants and therefore had substantially less power. If the α level for inclusion of predictors in each block is increased to 0.07, then DAT: Space Relations does enter the regression in block 2. In that case, IQ, Space Relations, and Mastery Motivation enter the regression overall and account for 85.5% of the variance among the participants' mean error, $R = 0.925$, $F(3,11) = 15.75$, $p = .001$.

2.3.4.2. Discussion. Important additional information about the nature of spatial reasoning and learning came from the examination of individual differences. There was very substantial variation among the performance of participants in this experiment, and 61.5% of this variation could be accounted for with three variables: (a) how much time people were willing to spend relative to their level of error, (b) their ability to manipulate spatial relations, and (c) their general capacity for fluid reasoning. Clearly, learning through the accumulation of intuitions was not in this case a passive reception of perceptual information. Quite the opposite was true. Participants succeeded in proportion to their ability to work hard at reasoning about spatial transformation.

2.3.5. Physical model

Mean error and response time in the test of generalization of reasoning to the physical model are presented in Fig. 12, broken down by whether or not the rotations were double-oblique and by group (i.e., the control group and each of the three VR learning conditions). The control group replicated standard findings. Thus, the mean error for the challenging rotations in that group was 57.1°, compared to 16.0° for the other rotations, $t(11) = 6.7$, $p < .001$. Mean response time was 30.5 and 11.2 s, respectively, $t(11) = 4.1$, $p = .002$.

There was substantial generalization from all three VR groups to performance with the physical model, as indicated by the lower error for the challenging rotations for the VR groups compared to the control group. The reliability of these differences were tested with Dunnett's C, a test of post hoc comparisons in which experimental groups are tested against a control condition (Visualization-Always, $p = .048$; Standard VR, $p = .002$; Visualization-Instruction, $p = .001$). In a within-VR test of generalization for the challenging rotations, a correlation of individual mean error

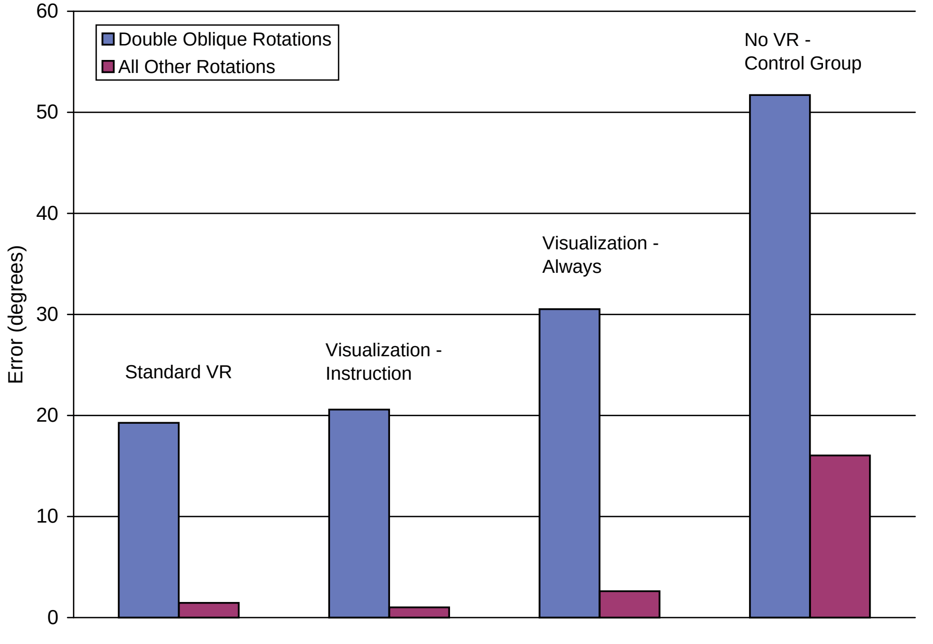

Fig. 12. Mean error and response time in reasoning about rotations of the physical model for the three VR learning conditions and the control group.

in the last 8 trial blocks in VR with mean error on the physical model was 0.657, $r^2 = 0.432$, $p < .01$.

Experience in VR reduced error across all types of rotation. When the four groups were compared (ANOVA) for both the easier and the more difficult rotations, there were main effects of group, $F(1, 44) = 99.9$, $p < .001$, and of type of rotation, $F(1, 44) = 113.8$, $p < .001$. Dunnett's C applied to error across all rotation types verified that there were differences between the VR groups when all rotations were considered together (Visualization-Always, $p = .002$; Standard VR and Visualization-Instruction, $p < .001$). The gain from experience in VR was substantially larger for the more difficult rotations (twice as large in the case of Standard VR and Visualization-Instruction), as indicated by an interaction between group and level of difficulty, $F(3, 44) = 3.02$, $p < .05$.

Overall, the common concern that experience in VR might not generalize to reasoning with real objects was addressed in this study. It was found that experience in VR did generalize to reasoning with real objects.

3. Experiment 2

The purpose of Experiment 2 was to clarify the nature of the information that participants were using to learn. It was quite possible, for example, that repetitive experience with the spatial reasoning problems was sufficient for learning in Experiment 1. After all, participants were generally correct in their reasoning when the axis of rotation was vertical or when the object was aligned with the axis of rotation. In addition, they often reported that the double-oblique rotations were challenging, a meta-cognitive observation that was consistent with higher response times for those problems. If the participants combined their knowledge of the problems that they were answering correctly and applied it to the problems they considered challenging, they might have generated a correct understanding of these also. For example, they might have inferred that oblique dishes will rotate around oblique axes in the same way that they rotate around vertical ones (see Pani & Dupree, 1994).

When the reasoning task included feedback and demonstration of the correct answers, there was repetition of the task across the domain of rotations along with two additional advantages: participants were given definite knowledge of errors, and they were provided substantial clues for thinking about how to generate correct answers. It was possible that participants already possessed critical components of knowledge necessary for correct reasoning about all of the rotations, but that they required explicit feedback and alternative answers to prompt and guide a reorganization of this knowledge.

Control and observation of the rotations after reasoning about them provided both repetition of the task and implicit feedback and demonstration of the correct answers. Control/observation added perception of the concrete details of the continuous motions. If it was important for the participants to assimilate new perceptual information about the motions, control/observation would improve learning beyond repetition and feedback/demonstration.

Overall, there was potential value for learning in repetition, feedback/demonstration, and control/observation, with each of the latter two encompassing the earlier one(s) and adding something to it. To provide a comparison among these learning situations, Experiment 2 was organized as a two-way between-groups factorial design. Whether or not participants received feedback and demonstration of the correct answers was one factor, and whether or not participants controlled and observed the motions after reasoning was the second. Of the four conditions generated by this combination, one group only solved the problems (Reasoning Only), one solved the problems and received feedback and demonstration of the answers (Feedback Only), one solved the problems and then controlled and observed the motions (Simulation Only), and one solved the problems, received feedback and demonstration, and controlled and observed the motions (Standard VR, as in Experiment 1).

3.1. Method and procedure

3.1.1. Participants

Participants were 48 university students, including 16 women and 32 men. They were paid \$7.50/h for their participation. Recruitment and screening of participants continued the methods of Experiment 1.

3.1.2. Design and materials

3.1.2.1. Psychometric tests. Each participant completed a series of psychometric tests prior to commencing the main experiment. These included the DAT: Space Relations, as a test of ability in spatial reasoning, and the short form of the WASI test of IQ. The WASI included the Matrices and the Vocabulary subtests. Assignment of participants to the four groups was balanced according to three criteria of decreasing priority: DAT: Space Relations, overall IQ, and gender.

The experiment was modeled on the Standard VR condition of Experiment 1 except for the changes necessary to generate the four conditions in which feedback/demonstration and control/observation were varied. Thus, in the Reasoning Only condition, the pause phase at the end of a trial came directly after the participant indicated that he or she had reached a final answer in the reasoning problem. Participants proceeded to the next trial when they were ready. In the Feedback Only condition, written feedback and a demonstration of the correct answer followed the reasoning phase of the trial, as in Experiment 1. The pause screen followed the feedback/demonstration. In the Simulation Only condition, the opportunity to control and observe the motion followed immediately after the reasoning phase of the trial. After the rotation was explored, the pause screen and then the next problem followed. In the Standard VR condition, feedback/demonstration followed problem solution and then control/observation followed the feedback (as in Experiment 1).

Instruction and practice with the VR controls were the same as in Experiment 1 except for changes that tailored the instruction to each of the four conditions. Each trial block again included 32 trials with a systematic variety of rotations. Because asymptotic performance was reached in Experiment 1 by 24 trial blocks, participants in Experiment 2 completed only 24 trial blocks (768 trials). Once again, participants were tested for a maximum of one hour per day.

3.2. Results and discussion

3.2.1. Comparisons among the four learning conditions

3.2.1.1. Results. Error in each trial was calculated as the shortest angle between the participant's setting of the blue response arrow and the correct answer. The primary measure of response time was the time taken to complete the reasoning phase of a trial (only that phase was included in all conditions). Performance with the double-oblique rotations across the four experimental conditions was of primary importance in this experiment. Mean error and mean response time for those rotations are presented in Fig. 13, broken down by the four experimental conditions and trial block.

Considering just the challenging rotations, there was a decrease in error across trial block for each of the four experimental conditions, including the Reasoning Only condition, $F(12.8,141) = 3.94$, $p < .001$ (for the four conditions together, $F(9.7,425) = 25.6$, $p < .001$). Hence, continuing to reason about a range of rotation problems did lead to an improvement in performance for the more challenging ones, even in the absence of explicit feedback or observation of the motions.

No comparisons of performance between the experimental conditions were statistically significant, and this raised an issue that had not been of practical concern in Experiment 1. That was that there were numerous participants who reasoned about rotation extremely well at the beginning of the experiment and therefore had little or nothing to learn. In an experiment designed to investigate influences on learning, such participants inappropriately inflate within-group variance and dilute effect sizes. The separation between those participants who might benefit from learning and those who have no need to can best be seen in the graph of individual performance for the condition in which learning was least evident. This was the Reasoning Only condition. Mean error per trial block for the challenging rotations in the Reasoning Only condition are presented in Fig. 14 for each of the participants. There was clearly a bimodal distribution that included participants who performed at a very high rate of accuracy throughout the experiment. (This was a larger problem in Experiment 2 than it had been in Experiment 1, apparently due to differences in who volunteered to serve in the experiments. Experiment 2 included many participants from the School of Engineering.)

To test for differences among the four learning conditions for those participants who could benefit from learning, participants in each condition who had the lowest mean error (across the entire experiment) for the challenging rotations were removed from the analyses. With anywhere from two to six participants excluded from each condition, statistical analyses suggested that participants who controlled and observed the motion after reasoning performed substantially better than those participants who did not (e.g., with four participants removed, $F(1,28) = 5.4$, $p = .028$). There was not an overall effect of whether feedback and demonstration of the correct answer were available, $F < 1.0$. However, there was a multivariate three-way

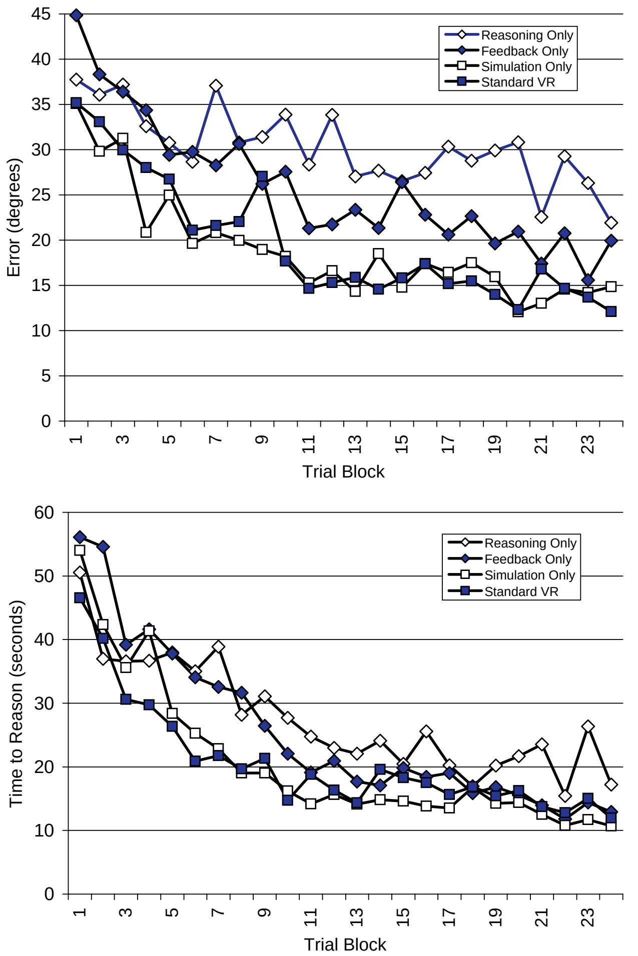

Fig. 13. Mean error and mean reasoning time for the double-oblique rotations in Experiment 2, broken down by trial block and the four experimental conditions.

interaction of control/observation with feedback/demonstration and trial block, $F(6, 23) = 4.98$, $p = .027$. The pattern of means illustrated in Fig. 13 clarifies the nature of this interaction: while there was a main effect of control/observation, and no main effect of feedback/demonstration, feedback/demonstration was better than just solving the problem repetitively, and this pattern emerged as learning progressed.

In all further statistical analyses, there were only minor differences between the data from the full samples of participants and the data from the samples with the more expert participants removed. For sake of simplicity, statistical results will be reported only for the full samples, unless otherwise indicated.

The amount of time spent reasoning about the rotation problems decreased over trials, $F(5.6, 1012) = 63.4$, $p < .001$. Participants who controlled and observed the

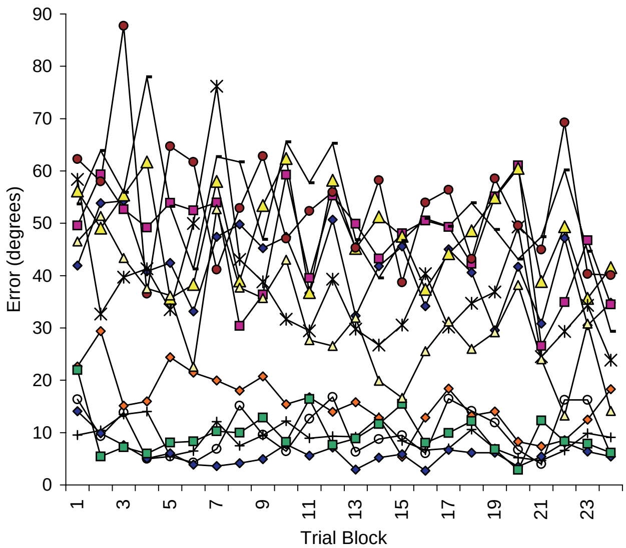

Fig. 14. Mean error for the double-oblique rotations in the Reasoning Only condition, broken down by trial block and individual participant.

motions were faster overall than those who did not, $F(1, 44) = 6.0$, $p = .018$. The effect of control and observation on the time to reason increased with practice, $F(6.8, 748) = 2.63$, $p = .013$.

A separate set of analyses examined differences in performance for the three easier types of rotation among the four experimental groups. No statistically significant differences were found among the groups for these rotations, either for error or response time.

3.2.1.2. Discussion. Learning to reason about rotations benefited from repetition of the variety of rotation problems, but it benefited more when repetition was coupled with additional information: either feedback and demonstration of the correct answers or the chance to control and observe the actual motions. Controlling and observing the motions clearly was the most effective way to learn. Moreover, if control and observation were available, adding explicit feedback and demonstration of the correct answer did not improve learning further.

3.2.2. Performance within the four learning conditions

When the four rotation types were compared within each of the four conditions, basic results from Experiment 1 were replicated. Given the main effect of control/observation, and the fact that performance in the two control/observation conditions was virtually identical, these two conditions were collapsed. In this case, a general improvement in error due to experience was indicated by a main effect of trial block, $F(6.5, 150) = 10.3$, $p < .001$. The double-oblique rotations were substantially more difficult than the other rotations, as indicated by the interaction between orientation of the dish to the shaft and orientation of the shaft in the environment, $F(1,23) = 17.8$, $p < .001$. There was convergence of performance for all types of rotation onto a high level of accuracy, indicated by an interaction of trial block with the two orientation parameters, $F(4.9,111) = 2.32$, $p = .05$. When the two feedback/demonstration conditions were collapsed, the same set of comparisons were statistically significant (trials, $F(3.2,74) = 7.7$, $p < .001$; dish by shaft orientation, $F(1,23) = 27.1$, $p < .001$; dish by shaft orientation by trials, $F(14.5,333) = 2.55$, $p = .002$).

3.2.3. Three models of spatial learning

3.2.3.1. Patterns of change across trials. In considering the three models of learning, a graphical examination of individual participant data suggested again that the intuition model was correct. Performance of individual participants characteristically switched between correct performance and large errors, with the proportion of correct responses increasing over time. A direct quantitative comparison of the three learning models again focused on predictions of the models for the change in error across blocks of trials. Mathematical simulations of error performance were constructed in accordance with each learning model for each of the four experimental conditions. The models were the same as those used earlier, except that the parts of the models that depended on the power function describing the data were fit separately to the power function in each of the four experimental conditions.

For all four conditions, and for either the entire sample of participants or the sample with the best participants removed, the intuition model was a very good fit to the data. The insight and spatial tuning models were poor fits. As in Experiment 1, the quality of the fit was tested statistically by converting the positive and negative derivatives of error across trials to absolute values and comparing the mean values from the simulated data to the mean values from the actual data with the Tukey HSD. For all four experimental conditions, the experimental data were not statistically different from the simulated data generated by the intuition model, $p > .35$. The experimental data were statistically different from the simulated data generated by the insight and spatial tuning models, $p < .02$. The simulated data from these two models did not differ, $p > .07$.

3.2.3.2. Item effects. As noted earlier, the intuition model is consistent with the presence of item effects. When the four experimental conditions were combined, item effects were identical to those in Experiment 1. These included effects of four display variables: the amount of rotation (with 180° more difficult than 90°; $r = .281$), whether or not the dish began the rotation in a Cartesian orientation with respect to the environment ($r = .238$), whether the dish faced up or down the shaft (with up being easier; $r = .122$), and whether the dish was partially or fully oblique in the environment (with partially oblique being easier; $r = .102$), $F(5,191) = 52.1$, $p < .001$.

When item effects were tested in individual conditions, their presence depended on the condition. The Reasoning Only condition generated the now standard set of item effects, including whether the dish began the rotation in a Cartesian orientation to the environment ($r = .429$), the amount of rotation ($r = .375$), whether the shaft was partially or fully oblique in the environment ($r = .125$), and whether the dish faced up or down the shaft ($r = .107$), $F(5, 191) = 27.2$, $p < .001$. Interestingly, the first two of these item effects were stronger than the effect of trials ($r = -.298$).

The Feedback Only condition generated three item effects. These included the amount of rotation ($r = .182$), whether the dish faced up or down the shaft ($r = .147$), and whether the dish began the rotation in a canonical orientation in the environment ($r = .159$), $F(4, 191) = 45.9$, $p < .001$. In this condition, none of the item effects was stronger than the effect of trials ($r = -.655$).

For the Simulation Only condition, there were no item effects. For the Standard VR condition, there were two item effects, amount of rotation ($r = .259$) and whether the shaft was partially or fully oblique ($r = .115$), $F(3, 191) = 65.0$, $p < .001$. When the Simulation Only and Standard VR conditions were combined, there was one item effect, the amount of rotation ($r = 0.189$), $F(2, 191) = 185.5$, $p < .001$.

In part, these tests for item effects replicated the results of Experiment 1 and provided further support for the intuition model. Certainly this was true for the combination of all four conditions into a single sample. On the other hand, the association of item effects with the weaker learning conditions suggests that effective learning overcomes them. The greater strength of item effects in Experiment 1 than in Experiment 2 for comparable conditions is a further demonstration that the participants in Experiment 2 were in general more skilled in reasoning about the rotations (and were introducing ceiling effects into the analyses of error).

3.2.3.3. Individual differences in the patterns of error. In this experiment, 15 out of 48 participants had mean error for the challenging rotations below 25° in the first two blocks of trials (which encompassed 16 trials with those rotations). Once again, we checked whether any of these participants formed domain-general descriptions of rotation consistent with the insight model. In particular, the raw error data from these participants were examined trial by trial for the challenging rotations in the early trial blocks. If insightful learning had occurred, there should have been instances in which error was high during the early trials (e.g., as high as 45°), but once a low level of error was reached (e.g., 25°.), no further large errors were committed for an extended period of time (e.g., 10 trials). Although this pattern did not occur at all in Experiment 1, the situation was more mixed in Experiment 2. There were six participants who never made an error as large as 45°. Another five participants made 1 error as large as 45°, but it did not occur at the beginning of the trials. There were six participants who had more than one error as large as 45°, but they were interspersed across several trials. Two participants made large errors early and then did not make any more. Certainly an absence of error, a single error, and errors only early in the trials could be said to reflect an understanding of rotation that was general across the domain. However, only two participants clearly fit the pattern of insightful learning in which early error changes suddenly to consistently correct performance.

3.2.4. Individual differences

Individual differences in the ability to reason about rotations were large, and there were strong relations between measures of individual differences and performance in the reasoning task. Given the main effect of control/observation, individual differences were examined separately for the half of the participants that did have control/observation and the half that did not.

Correlation matrices relating the psychometric measures and Mastery Motivation to mean error on the challenging rotations are presented in Table 3 for the two conditions that included control and observation of the motions and in Table 4 for the two conditions that did not. Overall, significant correlations spanned the types of measure, including general intelligence, spatial reasoning, and Mastery Motivation. The single highest correlation was for Mastery Motivation, which accounted for over 50% of the variance among the participants' mean error. The predictors were quite often correlated with each other.

The same method of hierarchical multiple regression was used to analyze the data from Experiment 2 as was used for Experiment 1. When accounting for mean error in the two conditions that included control/observation, Matrices entered the stepwise regression in the first block. This measure accounted for 23.3% of the variance among the participants' mean error, $R = 0.483$, $F(1,22) = 6.69$, $p = .017$. In the

Table 3
Correlation matrix (Pearson) relating error on the double-oblique rotations to the psychometric measures and mastery motivation for the Standard VR and Observe/Control conditions in Experiment 2

	Vocab	Matrices	IQ-percent	DAT:Space	MasMot
Error	$^{*}-0.360$	$^{**}-0.483$	$^{**}-0.459$	$^{***}-0.693$	$^{***}-0.742$
Vocab		$^{*}0.405$	$^{***}0.853$	$^{*}0.416$	$^{*}0.349$
Matrices			$^{***}0.732$	0.174	$^{*}0.366$
IQ-percent				0.277	$^{*}0.437$
DAT:Space					$^{*}0.372$

Individual entries correspond to mean error, WASI: Vocabulary, WASI: Matrices, WASI-IQ percentile, DAT: Space Relations, and Mastery Motivation. Asterisks indicate statistical significance (uncorrected) with * for $p < .05$, ** for $p < .01$, and *** for $p < .001$. $N = 24$.

Table 4
Correlation matrix (Pearson) relating error on the double-oblique rotations to the psychometric measures and mastery motivation for the Reasoning-Only and Feedback/Answer conditions in Experiment 2

	Vocab	Matrices	IQ-percent	DAT: Space	MasMot
Error	0.067	$^{***}-0.626$	-0.219	$^{*}-0.408$	$^{***}-0.842$
Vocab		-0.104	$^{***}0.763$	0.012	-0.026
Matrices			$^{**}0.483$	0.149	$^{**}0.511$
IQ-percent				0.125	0.158
DAT:Space					$^{**}0.505$

Individual entries correspond to mean error, WASI: Vocabulary, WASI: Matrices, WASI-IQ percentile, DAT: Space Relations, and Mastery Motivation. Asterisks indicate statistical significance (uncorrected) with * for $p < .05$, ** for $p < .01$, and *** for $p < .001$. $N = 24$.

second block of the regression, Space Relations entered. This measure accounted for an additional 38.3% of the variance, $R = 0.785$, $F(1,21) = 20.92$, $p < .001$. Together these two psychometric variables accounted for 61.6% of the variance among the participants on mean error. In the third block of the multiple regression, Mastery Motivation accounted for an additional 18% of the variance among the means, $R = 0.892$, $F(1,20) = 17.58$, $p < .001$. Altogether, these three variables accounted for 79.6% of the variance among the participants on mean error, $F(3,23) = 25.95$, $p < .001$.

A similar pattern of results occurred in the attempt to account for mean error in the Feedback Only and Reasoning Only conditions. Matrices entered the stepwise regression in the first block. This measure accounted for 39.2% of the variance among the participants' mean error, $R = 0.626$, $F(1,22) = 14.19$, $p = .001$. In the second block of the regression, Space Relations entered. This measure accounted for an additional 10.1% of the variance, $R = 0.702$, $F(1,21) = 4.2$, $p = .054$. Together these two psychometric variables accounted for 49.3% of the variance among the participants on mean error. In the third block of the multiple regression, Mastery Motivation accounted for an additional 26.8% of the variance among the means, $R = 0.873$, $F(1,20) = 22.45$, $p < .001$. Altogether, these three variables accounted for 76.1% of the variance among the participants on mean error, $F(3,23) = 21.27$, $p < .001$.

Overall, the same pattern of individual differences was observed for Experiment 2 as was observed for the Standard VR and Visualization-Instruction conditions of Experiment 1. In particular, quite large amounts of variance in individual performance were accounted for by Matrices, Space Relations, and Mastery Motivation.

4. General discussion

Rotational motion comprises an elegant domain of physical transformation. Points on a rotating object trace circular motions in parallel planes centered on a straight line, the axis of rotation. If the object is an orientable structure (e.g., with a definite top), the slant of the object is invariant to the axis of rotation, and the slant-direction of the object circles continuously around the axis. This single kinematic structure can join any two orientations in three-dimensional space, as Euler proved more than a century ago. Euler, however, was a leading mathematician, and following his reasoning requires formal training in mathematics. The difficulties that most people encounter in predicting the outcomes of double-oblique rotations indicate that Euler's Theorem is a non-intuitive description of how things move.

Double-oblique rotations are difficult to see or to reason about due to the relations between the rotating object and the spatial reference systems for orientation that are used in everyday perception and visual reasoning. In a double-oblique rotation, the axis and planes of motion are aligned neither with the intrinsic structure of the object nor with the principal frame of the environment. As a consequence, the rotational structure of the motion is not salient. At the same time, the person is disposed to perceive the orientation of the object relative to the environment (rather than the axis and planes of rotation). The outcome is perception of a continuous change of object orientation but not of a simple rotation. If a person reasons about the outcome of a double-oblique rotation, they are likely to visualize a motion aligned with the object or the environment, and their reasoning will be incorrect.

4.1. Spatial organization in spatial reasoning

Such phenomena indicate that perception, reasoning, and mental imagery about elementary physical properties are not necessarily composed in a system that enforces the invariant properties of physical space. Ordinary visual reasoning is based on a set of mental models constrained by spatial organization acquired in interaction with the physical world (Clement, 1983; McCloskey, 1983; Pani, 1997, 1999; Tversky, 1981). This reasoning is sophisticated, but it does not incorporate universal analytical categories or a unified system of logical inference. Indeed, in numerous physical domains, people may reason in ways that generate clear inconsistencies between their answers and the premises of the problem (e.g., indicating a "cube" with six corners; Hinton, 1979; Pani et al., 1997).

The differences between the easier rotations studied here and the more difficult ones are examples of structural distinctions that have broad application in spatial reasoning. A consideration of these structural distinctions clarifies how processes of learning interact with spatial domain knowledge across a large set of physical structures.

4.1.1. Some foundations of useful structure

4.1.1.1. Alignment. The rotations that are easier to understand include alignment of the motion, in the form of parallel or perpendicular orientations, to salient spatial reference systems. Rotation about an axis that is vertical, or that is aligned with an obvious axis of the object, is relatively easy to see and to reason about. The immediate grasp of spatial relations that include alignment is common across the domain of physical reasoning. If people are asked to predict the outcomes of projective transformations, for example, such as to predict the shape of the shadow that an object will cast, they are substantially more skilled when the intrinsic structure of the object is aligned with the direction of projection (Pani, Jeffres, Shippey, & Schwartz, 1996). Across spatial domains, the critical structures and the effective reference systems may vary, but the value of alignment appears to be universal (e.g., Pani, 1997, 1999; Pani, William, & Shippey, 1998; Shelton & McNamara, 2001).

4.1.1.2. Symmetry. Alignment of a structure with a spatial reference system is an instance of symmetry in the general sense of the term—a spatiotemporal property that is invariant despite a transformation of the object. It is conventional to define symmetries in terms of the transformations that leave the object invariant. Thus, there is translational symmetry (e.g., in a picket fence), rotational symmetry (in a flower), reflection symmetry (in a face), and so on (e.g., Cederberg, 1989). Two parallel orientations, such as a vertical shaft and a vertical wall, have translational symmetry, and a perpendicular orientation, such as a vertical shaft in relation to the ground, has reflection and (quite often) rotational symmetry.

All symmetries have an inherent redundancy associated with the transformation that defines them. For example, a translational symmetry is redundant as one translates across the object. As a consequence, a representational system that is sensitive to symmetry will be able to form more efficient descriptions when there is symmetry in a scene (e.g., Attneave, 1954; Garner, 1974). If a perceived environment has a salient vertical, for example, then perception of an object that is vertical does not add a new orientation. Rather, it is one more instance in which the environment includes a vertical. In like manner, placing a series of objects at a single orientation is much like making them all one color. They are seen as a series of objects with a single orientation (consider Beck, 1966).

The efficiency of symmetry has two important implications for physical reasoning. First, reasoning about physical properties related to orientation can proceed with a relatively simple description of the scene (also see Proffitt & Gilden, 1989). Second, orientations aligned with a salient reference system themselves become salient. Thus, an axis of rotation that is vertical has an orientation that is already in the scene, and it becomes a salient reference axis for perception of the orientation of an object that will rotate around it. The orientation of the rotating object, moreover, will have the same orientation to the axis of rotation as it does to the vertical, further simplifying description of the scene.

4.1.1.3. Singularity. Symmetries are examples of spatial properties that are singular. Singular properties arise as categorically unique values along dimensions of variation (Goldmeier, 1972, 1982; Leyton, 1992; Pani, 1997; Rock, 1983; Wertheimer, 1950/1923). Examples of singular values include "maximum," "minimum," "same," and "orthogonal." Singular spatial relations include "straight," "circular," "vertical," "horizontal," "midpoint," and "center."

Singular terms are qualitative (i.e., nonmetric). We do not need a ruler to identify the object that is longest or shortest, or to find two things that are the same. Parallel and perpendicular are single angles, of 0° and 90°, but they form qualitative distinctions that comprise categories at the same level as acute and obtuse (in which there are infinite numbers of angles). This aspect of singularities makes them distinctive. For example, when an object faces straight out from an axis of rotation, it has an easily recognized orientation and one that is easy to visualize as the individual imagines the object to rotate about the axis.

4.1.1.4. Affordances of alignment, symmetry, and singularity. Reliance on alignment, symmetry, and singularity will continue only so long as it is adaptive. In fact, these properties are quite valuable in the search for useful information. Singular properties are nonarbitrary values—sometimes the only nonarbitrary values—that are available for making comparative judgments. In measuring objects, we measure them across their longest or shortest extent. In judging tradeoffs, we look for a balance. We fill the gas tank to the top and take the shortest path to our destination. In the case of symmetry, building simple rotational systems with alignment between objects and axes of rotation has been extremely useful in human engineering, providing wheels, pulleys, knobs, and cranks.

More generally, the physical relations of balance, alignment, and "fitting together" are the basis of innumerable useful physical systems. Putting a round peg in a round hole is an instance of alignment and translational symmetry, as is packing the back of a truck with rectangular boxes. Even in those situations in which it would seem that there is substantial choice, people employ alignment, symmetry, and singularity. For example, it is easy to tell which state borders in the United States were inherited from nature as rivers, coastlines, or mountain ranges, and which borders were generated from human considerations. Natural borders tend to be irregular, but human ones tend to be straight and parallel, typically running north/south or east/west.

Altogether, a cognitive system that is sensitive to the presence of alignment, symmetry, and singularity is tuned to simplicity, distinctiveness, and pragmatically useful descriptions of physical relations. When such a system encounters physical events that contain these properties, it will find them to be readily comprehensible. As a consequence, rotational motions aligned with the intrinsic structure of the object or the vertical of the environment are readily visualized by most people without instruction.

4.1.1.5. Unfamiliar structures. There are many motions in the environment that we see as arbitrary, complicated, or "interesting," including falling leaves, hand gestures, tumbling boxes, and the gyrations of Olympic divers. Double-oblique rotations are quite common, but due to our preferences for organizing orientation, and to the fact that the rotational properties of tumbling boxes are not of great concern, we do not organize them as simple rotations. Similarly, shadows come in every variety (e.g., as the sun moves across the sky), but we do not generally think about the geometric relations that generated them. Variations of phenomena that do not possess alignment, symmetry, or singularity may benefit only from minimal levels of spatial organization (e.g., as a continuous motion rather than a rotation).

4.2. Spatial learning

The primary question addressed by the research reported here was how learning to reason correctly across the domain of simple rotations would progress. Would learning be efficient for people not trained in formal methods? Would learning lead to insights into the kinematics of rotation? If not, would learning take place through the accumulation of intuitions about rotations that were seen? We have characterized intuitions as knowledge that encodes information from relatively specific classes of perceptual experience. Alternatively, would there be a process of spatial tuning of responses?

In two longitudinal experiments, people learned without formal training in geometry or kinematics to reason about double-oblique rotations. Some people learned quickly—some so quickly that long response times in early trials were the only evidence that they needed to learn at all. Others ended the experiment still committing relatively large errors, but with far less error than at the beginning. In looking across individuals, it was clear that learning to reason about double-oblique rotations was a demanding task. Mean response time was long, and the more time that participants were willing to give to the task, relative to their level of error, the better they performed. Indeed, mastery motivation was the single best predictor of individual differences in performance.

The pattern of performance across these two experiments clearly supported the view that learning to reason across the domain of simple rotations took place as the accumulation of spatial intuitions. As the process of learning continued through multiple trials, a standard pattern was for correct answers to be followed by large errors. An incremental decrease in mean error actually was the product of an increasing proportion of accurate answers. It appears that participants learned how rotations took place for certain configurations of the assembly and then were able to predict correctly the outcomes of rotations for similar configurations. When the assembly looked substantially different from the configurations that had been learned, generalization often failed. The new situation was then studied and learned. As learning progressed in this manner, some participants did show sudden improvements in overall performance, suggesting that relatively general schemas were formed (see Fig. 7). Numerous indicators supported the validity of the intuition model, including graphical presentation of the raw data, statistical analysis and computer simulation of the pattern of change in performance across trials, multiple item effects, and the participants' common assertion that they solved the problems by remembering ones that they had seen before (although they had not seen identical problems).

Additional results of these experiments shed light on the nature of the process of developing intuitions in this domain of physical reasoning. In one of the conditions of Experiment 1, computer graphics helped participants to provide correct answers for more than 100 trials with the difficult rotations, and in every case these individuals controlled and observed the motion afterward. These participants no doubt became familiar with the physical structures and motions in the scene. However, the participants were unable later to recall the scene in a way that could support correct reasoning. Clearly these individuals were not attending to the spatiotemporal relations among the elements of the problems that would have allowed identifying initial configurations of the physical assembly and associating them with the outcomes of the rotations (see also Goldstone, 1998; Goldstone & Stevyers, 2001; Schyns, Goldstone, & Thibaut, 1998).

The pattern of results for the psychometric predictors of performance makes a complementary point. Participants who showed greater skill in fluid reasoning performed better overall. The matrices tasks that provide measures of fluid reasoning are tasks that require a goal-directed discovery of relations that characterize a collection of objects. In addition, participants who entered the experiment with more skill in thinking about spatial transformation performed better overall. Together, these results suggest the importance of an ability to discover and encode informative properties of the physical assembly and to integrate this information into reliable event level knowledge of the motions.

Skill in reasoning advanced the ability to understand individual types of problem without guaranteeing that those problems would be understood to be part of a broad class. In other words, the research team knew that there was a single class of problems that could be described as double-oblique rotations. However, when individual participants developed an effective way to think about a particular rotation, they did not necessarily understand that it was an example of a broad class. Rather, a variety of implicit subcategories of the rotations were initially formed. Skill in reasoning was not expressed in a single "Aha!" so much as it was by a series of "Hmmm's" and "Ahhh's" that accumulated across the domain.

When different learning situations were compared in Experiment 2, it was found that repeated experience with reasoning across the domain of rotations was better than nothing. Repetition presumably allowed people to compare their answers to problems that they found easy with their answers to more challenging problems. Learning was substantially enhanced when participants also were given immediate feedback about their answers along with demonstrations of the correct answers. The best aid to learning, however, was in controlling and observing the actual motions. That is, the most effective way to learn went beyond a reorganization of existing knowledge and incorporated new perceptuomotor experience with the motions. It seems likely to us that this form of learning included the encoding of intrinsically spatiotemporal information about the continuous paths of the motions. While this knowledge would be primarily nonverbal, it might be expressed in such statements as, "When the dish points up the shaft like that, the rotation will trace a circular path around the top of the shaft." Such spatiotemporal representation would be one format for integrating informative properties of the rotations into event level intuitions.

The conservative aspect of spatial intuitions with regard to generalization across a domain need not be a disadvantage overall. To the contrary, it may be an advantageous compromise, perhaps an ideal one, for a cognitive system that must reason and predict as well as learn and recognize. Systems that are highly powerful in their reasoning may generate too many alternative outcomes for efficient evaluation, and many of these generated outcomes may be patently false (Forbus, 2001). Conditioning the growth of categories and processes of inference on concrete experience provides a conservatism that keeps reasoning relevant and efficient (Brooks, 1987; Forbus & Gentner, 1997; Forbus, 2001; Medin & Ross, 1989). If more abstract understanding in a domain is based on well-developed intuitions, a great deal of time and error will have been saved in the long run.

4.3. Conclusion

Spatial learning in the situations that we have studied is a determined effort to move outside the boundaries of the typical system for organizing spatiotemporal relations. It involves resisting the default organization of a physical assembly and its motion and forming a new system that can be used to accurately predict outcomes. The result is that motions that had been perceived as "things moving around" can instead be understood to be systematic and predictable.

Here we have found that people do not proceed by forming classical insights—understanding which, once achieved, is domain-general. Nor is it a continuous reduction of error that gradually settles onto correct answers. Considerable hard work forms intuitions—articulated event level knowledge that is tied to particular subclasses of perceptual experience. Extending this learning across a physical domain requires continuing to explore the domain and to acquire new intuitions. Ultimately these intuitions may be consolidated into schemas that are relatively general.

The acquisition of spatial intuitions is consistent with studies of expertise in which advanced practitioners have developed the ability to recognize useful configurations (e.g., Chase & Simon, 1973; Egan & Schwartz, 1979; Reitman, 1976). Here, we have seen this capability form the core of new learning in a fundamental physical domain. Moreover, it has developed in conjunction with an ability to anticipate the outcomes of spatial transformations.

The abstract and domain-general understanding that is embodied in mathematics and formal science comes from an overlapping, but ultimately different, set of representations (consider Chi et al., 1981). Perhaps it arises in part from further exploration and experience that combine and abstract what is known in intuitions and schemas. It should be remembered in this regard, however, that much of what is studied in the advanced physical curriculum, such as motion in the absence of gravity or friction, is not available in everyday experience. Partly as a consequence, abstract physical knowledge has depended for its development on the efforts of great thinkers such as Euler, Galileo, Newton, and Einstein. Part of their accomplishment has been to develop conceptual systems that appear to contradict everyday intuitions (Clement, 1983; McCloskey, 1983). Higher level learning through the further acquisition of intuitions, therefore, requires thought experiments or physical simulations that lay down a new series of intuitions to combine with the old (e.g., riding on a light beam). Indeed, enlivening new intuitions may be the best method for introducing advanced knowledge to most people, and graphical computer simulation is being deployed in this effort (e.g., Forbus, Ureel, Carney, & Sherin, 2004; Goldstone & Sakamoto, 2003; Miller, Lehman, & Koedinger, 1999). On the other hand, truly abstract domain-general physical knowledge is ultimately embodied in formal analytical systems that are encountered in classroom and technical settings. Discourse in the context of such systems encourages knowledge that is abstract and general in a way that is unlikely to arise through concrete experience alone.

References

Attneare, F. (1954). Some informational aspects of visual perception. *Psychological Review, 61*, 183–193.

Beck, J. (1966). Effects of orientation and shape similarity on perceptual grouping. *Perception & Psychophysics, 1*, 300–302.

Bennett, G. K., Seashore, H. G., & Wesman, A. G. (1989). *Differential aptitude tests for personnel and career assessment: Space relations*. San Antonio, TX: The Psychological Corporation, Harcourt Brace Jovanovich.

Brooks, L. R. (1987). Decentralized control of categorization: The role of prior processing episodes. In U. Neisser (Ed.), *Concepts and conceptual development* (pp. 141–174). Cambridge, UK: Cambridge University Press.

Brooks, L. R., Norman, G. R., & Allen, S. W. (1991). Role of specific similarity in a medical diagnostic task. *Journal of Experimental Psychology: General, 120*, 278–287.

Cederberg, J. N. (1989). *A course in modern geometries*. New York: Springer-Verlag.

Chase, W. G., & Simon, H. A. (1973). The mind's eye in chess. In W. G. Chase (Ed.), *Visual information processing*. New York: Academic Press.
Cheng, P. N., & Holyoak, K. J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology, 17*, 391–416.
Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*, 121–152.
Clement, J. (1983). A conceptual model discussed by Galileo and used intuitively by physics students. In D. Gentner & A. L. Stevens (Eds.), *Mental Models* (pp. 325–340). Hillsdale, NJ: Erlbaum.
Cohen, J., & Cohen, P. (1975). *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum.
Cooper, L. A., & Shepard, R. N. (1973). Chronometric studies of the rotation of mental images. In W. G. Chase (Ed.), *Visual information processing* (pp. 75–176). NY: Academic Press.
Egan, D., & Schwartz, B. (1979). Chunking in the recall of symbolic drawings. *Memory & Cognition, 7*, 149–158.
Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Manual for kit of factor-referenced cognitive tests*. Princeton, NJ: Educational Testing Service.
Forbus, K. D. (2001). Exploring analogy in the large. In D. Gentner, K. J. Holyoak, & B. N. Kokinov (Eds.), *The analogical mind: perspectives from cognitive science* (pp. 23–58). Cambridge, MA: MIT Press.
Forbus, K. D., Ureel, L. C., Carney, K., & Sherin, B. L. (2004). Qualitative modeling for middle-school students. In *Proceedings of the 18th international qualitative reasoning workshop* (pp. 81–87). USA, Evanston, Il.
Forbus, K. D., & Gentner, D. (1997). Qualitative mental models: Simulations or memories? In *Proceedings of the eleventh international workshop on qualitative reasoning* (pp. 99–104). Italy, Cortona.
Garner, W. R. (1974). *The processing of information and structure*. Potomac, MD: Erlbaum.
Gibson, J. J. (1950). The perception of visual surfaces. *American Journal of Psychology, 63*, 367–384.
Gick, M., & Holyoak, K. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15*, 1–38.
Gigerenzer, G., Todd, P. M., & The ABC Research Group. (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.
Goldmeier, E. (1972). Similarity in visually perceived forms. *Psychological Issues, 8*(1), 1–135 (Translation of a paper originally published in German in 1936).
Goldmeier, E. (1982). *The memory trace: Its formation and its fate*. Hillsdale, NJ: Erlbaum.
Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology, 49*, 585–612.
Goldstone, R. L., & Sakamoto, Y. (2003). The transfer of abstract principles governing complex adaptive systems. *Cognitive Psychology, 46*, 414–466.
Goldstone, R. L., & Stevyers, M. (2001). The sensitization and differentiation of dimensions during category learning. *Journal of Experimental Psychology: General, 130*, 116–139.
Hinton, G. (1979). Some demonstrations of the effects of structural descriptions in mental imagery. *Cognitive Science, 3*, 231–250.
Hogarth, R. M. (2001). *Educating intuition*. Chicago: The University of Chicago Press.
Holden, A. (1971). *Shapes, space, and symmetry*. New York: Columbia University Press.
Kaiser, M. K., Jonides, J., & Alexander, J. (1986). Intuitive reasoning about abstract and familiar physics problems. *Memory & Cognition, 14*, 308–312.
Larkin, J. H., McDermott, J., Simon, D. P., & Simon, H. A. (1980). Expert and novice performance in solving physics problems. *Science, 208*, 1335–1342.
Logan, G. D. (2002). An instance theory of attention and memory. *Psychological Review, 109*, 376–400.
Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge, MA: Harvard University Press.
Just, M. A., & Carpenter, P. A. (1985). Cognitive coordinate systems: Accounts of mental rotation and individual differences in spatial ability. *Psychological Review, 92*, 137–172.
Kohler, W. (1929). *Gestalt psychology*. New York: Liveright.
Leyton, M. (1992). *Symmetry, causality, mind*. Cambridge, MA: MIT Press.
Mackintosh, N. J. (1998). *IQ and human intelligence*. New York: Oxford University Press.

Massironi, M., & Luccio, R. (1989). Organizational versus geometric factors in mental rotation and folding tasks. *Perception, 18*, 321–332.

McCloskey, M. (1983). Naive theories of motion. In D. Gentner & A. L. Stevens (Eds.), *Mental models* (pp. 299–324). Hillsdale, NJ: Erlbaum.

Medin, D. L., & Ross, B. H. (1989). The specific character of abstract thought: Categorization, problem-solving, and induction. In R. J. Sternberg (Ed.). *Advances in the psychology of human intelligence* (Vol. 5, pp. 189–223). Hillsdale, NJ: Erlbaum.

Miller, C. S., Lehman, J. F., & Koedinger, K. R. (1999). Goals and learning in microworlds. *Cognitive Science, 23*, 305–336.

Nosofsky, R. M., & Johansen, M. K. (2000). Exemplar-based accounts of ''multiple-system'' phenomena in perceptual categorization. *Psychonomic Bulletin & Review, 7*, 375–402.

Pani, J. R. (1989). *Orientation of the object and slant of the axis of rotation in perception and mental imagery of object rotation*. Poster presented at the annual meeting of the Psychonomic Society, Atlanta, GA.

Pani, J. R. (1993). Limits on the comprehension of rotational motion: Mental imagery of rotations with oblique components. *Perception, 22*, 785–808.

Pani, J. R. (1997). Descriptions of orientation in physical reasoning. *Current Directions in Psychological Science, 6*, 121–126.

Pani, J. R. (1999). Descriptions of orientation and structure in perception and physical reasoning. In E. Winograd, R. Fivush, & W. Hirst (Eds.), *Ecological approaches to cognition: Essays in honor of Ulric Neisser* (pp. 59–95). Mahwah, NJ: Erlbaum.

Pani, J. R. (2000). Cognitive description and change blindness. In D. L. Simons (Ed.), *Visual cognition: Change blindness and visual memory* (pp. 107–126). East Sussex, UK: Psychology Press, [special issue].

Pani, J. R., Chariker, J. H., & Fell, R. D. (2005). Visual cognition in microscopy. Poster presented at the annual meeting of the Cognitive Science Society. Stresa, Italy. Abstract published in *Proceedings of the 27th annual conference of the cognitive science society*.

Pani, J. R., & Dupree, D. (1994). Spatial reference systems in the comprehension of rotational motion. *Perception, 23*, 929–946.

Pani, J. R., Jeffres, J. A., Shippey, G., & Schwartz, K. (1996). Imagining projective transformations: Aligned orientations in spatial organization. *Cognitive Psychology, 31*, 125–167.

Pani, J. R., William, C. T., & Shippey, G. (1995). Determinants of the perception of rotational motion: Orientation of the motion to the object and to the environment. *Journal of Experimental Psychology: Human Perception and Performance, 21*, 1441–1456.

Pani, J. R., William, C. T., & Shippey, G. T. (1998). Orientation in physical reasoning: Determining the edge that would be formed by two surfaces. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 283–300.

Pani, J. R., Zhou, H., & Friend, S. M. (1997). Perceiving and imagining Plato's solids: The generalized cylinder in spatial organization of 3D structures. *Visual Cognition, 4*, 225–264.

Parsons, L. M. (1987). Visual discrimination of abstract mirror-reflected three-dimensional objects at many orientations. *Perception and Psychophysics, 42*, 49–59.

Parsons, L. M. (1995). Inability to reason about an object's orientation using an axis and angle of rotation. *Journal of Experimental Psychology: Human Perception and Performance, 21*, 1259–1277.

Proffitt, D. R., & Gilden, D. L. (1989). Understanding natural dynamics. *Journal of Experimental Psychology: Human Perception and Performance, 15*, 384–393.

Proffitt, D. R., Kaiser, M. K., & Whelan, S. M. (1990). Understanding wheel dynamics. *Cognitive Psychology, 22*, 342–373.

Rock, I. (1983). *The logic of perception*. Cambridge, MA: Bradford/MIT Press.

Reitman, J. (1976). Skilled perception in Go: Deducing memory structures from inter-response times. *Cognitive Psychology, 8*, 336–356.

Rosch, E., & Mervis, C. B. (1975). Family resemblance studies in the internal structure of categories. *Cognitive Psychology, 7*, 573–605.

Ross, B. H. (1984). Remindings and their effects in learning a cognitive skill. *Cognitive Psychology, 16*, 371–416.

Schone, H. (1984). *Spatial orientation*. (C. Strausfeld, Trans.). Princeton, NJ: Princeton University Press (Original work published 1980).

Schyns, P. G., Goldstone, R. L., & Thibaut, J. (1998). Development of features in object concepts. *Behavioral and Brain Sciences, 21*, 1–54 (Target article).

Shelton, A. L., & McNamara, T. P. (2001). Systems of spatial reference in human memory. *Cognitive Psychology, 43*, 274–310.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171*, 701–703.

Shepard, R. N. (2001). Perceptual-cognitive universals as reflections of the world. *Behavior and Brain Sciences, 24*, 581–601.

Shiffrar, M. M., & Shepard, R. N. (1991). Comparison of cube rotations around axes inclined relative to the environment or to the cube. *Journal of Experimental Psychology: Human Perception and Performance, 17*, 44–54.

Simons, D. L. (Ed.). (2000). *Visual cognition: Change blindness and visual memory [special issue]*. East Sussex, UK: Psychology Press.

Tversky, B. (1981). Distortions in memory for maps. *Cognitive Psychology, 13*, 407–433.

Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence*. San Antonio, TX: The Psychology Corporation, Harcourt Brace Jovanovich.

Wertheimer, M., 1950. Laws of organization in perceptual forms. In W. D. Ellis (Ed. and Trans.), *A source book of Gestalt Psychology* (pp. 71–88). Humanities Press: New York (Abridged version of work originally published 1923).

Wertheimer, M. (1959). *Productive thinking*. New York: Harper.