Computer-Based Learning of Neuroanatomy: A Longitudinal Study of Learning, Transfer, and Retention

Julia H. Chariker, Farah Naaz, and John R. Pani
University of Louisville

A longitudinal experiment was conducted to evaluate the effectiveness of new methods for learning neuroanatomy with computer-based instruction. Using a three-dimensional graphical model of the human brain and sections derived from the model, tools for exploring neuroanatomy were developed to encourage *adaptive exploration*. This is an instructional method that incorporates graphical exploration in the context of repeated testing and feedback. With this approach, 72 participants learned either sectional anatomy alone or whole anatomy followed by sectional anatomy. Sectional anatomy was explored either with perceptually continuous navigation through the sections or with discrete navigation (as in the use of an anatomical atlas). Learning was measured longitudinally to a high performance criterion. Subsequent tests examined transfer of learning to the interpretation of biomedical images and long-term retention. There were several clear results of this study. On initial exposure to neuroanatomy, whole anatomy was learned more efficiently than sectional anatomy. After whole anatomy was mastered, learners demonstrated high levels of transfer of learning to sectional anatomy and from sectional anatomy to the interpretation of complex biomedical images. Learning whole anatomy prior to learning sectional anatomy led to substantially fewer errors overall than learning sectional anatomy alone. Use of continuous or discrete navigation through sectional anatomy made little difference to measured outcomes. Efficient learning, good long-term retention, and successful transfer to the interpretation of biomedical images indicated that computer-based learning using adaptive exploration can be a valuable tool in instruction of neuroanatomy and similar disciplines.

Keywords: learning, instruction, neuroanatomy, graphics, interactive

Anatomy is a challenging topic to learn for several reasons. Large volumes of material must be learned in relatively short periods of time. Anatomical structures may have irregular and indistinct shapes, they may have little variation in color and texture, and they may be related to each other in complex three-dimensional arrangements. Beyond this intrinsic challenge, comprehensive instruction in anatomy includes mastery of sectional anatomy, a representation in which anatomy is viewed in two-dimensional planar sections sampled from the three-dimensional tissue. Sectional presentation of anatomy is common in biomedicine (e.g., in medical imaging, light microscopy, histological prep-

arations, and dissection), and it is particularly difficult to master (Crowley, Naus, Stewart, & Friedman, 2003; Pani, Chariker, & Fell, 2005; see also Lesgold, Rubinson, Feltovich, Glaser, Klopfer, & Wang, 1988).

Sectional anatomy is challenging in part because a spatial transformation occurs when a two-dimensional section is taken from a three-dimensional object. The two and three-dimensional structures may look very different from each other. In addition, there is often not a one-to-one mapping between the representations of whole and sectioned anatomy. For example, anatomy can be sectioned at different depths and at a variety of planar orientations. Typically, this will generate substantial variation in the form of the presented anatomy, a one-to-many mapping. Many-to-one mappings also occur because differently shaped three-dimensional structures can appear similar in a sectional image (e.g., a small sphere and a long tube).

Compounding the challenge of anatomy learning, dissection labs are expensive to maintain, and training in such settings is time consuming. As a consequence, there is increasing interest in exploring new pedagogical capabilities made available by computer-based graphical approaches to anatomy education (Reeves, Aschenbrenner, Wordinger, Roque, & Sheedlo, 2004). Computers can present graphical material effectively, and they are appropriate for large domains in which extensive self-study is beneficial.

In the present study, we developed, tested, and compared several basic options for learning neuroanatomy in an interactive computer-based environment. Considering the common use of sectional imagery and the difficulty of mastering it, a primary goal

was to move toward an optimal computer-based system for learning sectional neuroanatomy. Our research questions centered on the idea that facilitating cognitive organization of the mass of information contained in sectional anatomy will increase the efficiency of learning.

The main hypothesis of this study was that prior knowledge of whole anatomy would serve as a relatively efficient basis for acquiring and organizing the information in sectional anatomy. Specifically, we expected learning of sectional anatomy to be more efficient if the corresponding whole anatomy had been learned first. In addition, we expected there to be no cost, and potentially some benefit, of this more efficient learning for long-term retention of sectional anatomy and for transfer of knowledge to the interpretation of novel anatomical representations (e.g., Visible Human and magnetic resonance imaging [MRI] images not seen previously).

From a practical point of view, anatomical atlases have established a standard for the presentation of sectional neuroanatomy that is highly systematic. The situation for whole neuroanatomy, however, is left to the preferences of each team of instructional designers. When a concerted effort is made to present whole anatomy, it is typically presented in terms of depictions of the entire brain as it can appear from the outside (Bear, Connors, & Paradiso, 2007; Duvernoy, 1999). Structures that are inside the brain are typically seen only in sectional representations. In general, the questions of whether and how whole and sectional anatomy should be integrated in an instructional system are not asked. Our intent is to address these questions in a systematic program of research.

A second hypothesis in this study was that providing smooth interactive movement through a continuous series of sectional representations also would facilitate organization of sectional anatomy. This is a presentation of sectional anatomy that has become common in recent computer-based anatomical representations (e.g., Hohne et al., 2003; Mai, Paxinos, & Voss, 2008). However, we do not know of any studies of its relative value for learning. An alternative method also presented sections in serial order, as that is typical in anatomical atlases. Navigation through the sections, however, was discrete, like paging through a paper atlas.

Tests of the two main hypotheses of this study were combined in a 2 × 2 factorial design. Half of the participants learned only sectional anatomy (the sections alone groups) and half of the participants learned whole anatomy followed by sectional anatomy (the whole plus section groups). In each of these cases, half of the participants learned sectional anatomy with a smooth, seamless, interactive presentation (the continuous groups), and the other half learned sectional anatomy with the discrete presentation (the discrete groups).

The computer-based systems described here are under development, and it would be premature to compare them to conventional instructional methods. Nonetheless, two tests that took place after learning was completed provided an absolute standard for an initial assessment of these systems. First, tests of long-term retention (after 2–3 weeks) provided a conventional evaluation that is measured on a ratio scale (percentage retained). Second, tests of the transfer of knowledge to the interpretation of biomedical images (Visible Human and MRI) provided a gold-standard test of the external validity of the instructional methods.

In the remainder of this introduction, we briefly describe the prior evidence in relation to the two main hypotheses for this study. We then describe the strategic approach that was taken in the development of the computer-based instructional systems and the evaluation of them.

Transfer of Learning
From Whole to Sectional Anatomy

Several lines of research support the hypothesis that learning whole anatomy will facilitate learning of sectional anatomy. This research has generally been conducted, however, with verbal, categorical, and logical material. It is often not clear how it can be generalized to domains with a significant spatial component.

Presenting information in an organized framework makes learning more efficient (Bower, Clark, Lesgold, & Winzenz, 1969; Bransford, Brown, & Cocking, 2000). Whole anatomy would seem to be an ideal framework for organizing the information contained in sectional representations that are derived from it. In addition, representation of a domain in more than one format supports robust learning (Paivio, 1969) and flexible reasoning (de Jong et al., 1998) and is associated with expertise (van Someren, Reimann, Boshuizen, & de Jong, 1998). Finally, knowledge of whole anatomy may serve as a mental model—a descriptive knowledge structure that serves as a concrete basis for inference (Byrne & Johnson-Laird, 1989; Hegarty, 1992; Schwartz & Heiser, 2006).

In the study of cognition related to biomedical imagery, verbal protocol studies of the identification of sectional anatomy in microscopy indicate that successful recognition depends on the use of general anatomical knowledge (Crowley et al., 2003; Pani, Chariker, & Fell, 2005). Similar findings occur in the literature on the interpretation of x-ray images (Lesgold et al., 1988; Nodine & Kundel, 1987).

A contrasting view is that the best method for learning to interpret specialized imagery is direct practice with correct labeling of the imagery. In support of this view, people often learn a domain in terms of individual cases rather than global models or principles (Brooks, Norman, & Allen, 1991; Kolodner & Leake, 1996; Pani, Chariker, Dawson, & Johnson, 2005). Whole and sectioned anatomical structures look quite different from each other, and it is not obvious that knowledge would transfer between instances of these representations (consider Ross, 1984). More generally, information that one might expect to transfer to a new problem or domain very often does not (Barnett & Ceci, 2002; Gick & Holyoak, 1980, 1983). Overall, it was an empirical question whether prior knowledge of whole anatomy would make learning sectional anatomy more efficient.

Presentation of Sectional Anatomy

Work in *anorthoscopic perception* and *kinetic completion* supports the hypothesis that continuous interactive navigation through sectional imagery would facilitate organization of the spatial information in the sections and improve learning (for a review of research on these topics, see Palmer, 2002). In general, the visual system is tuned to perceive unitary objects across sequences when appropriate information is made available. In anorthoscopic perception, for example, an object that is partially visible through a

small aperture is perceived accurately, so long as the structure of the object is revealed by motion of the object or of the aperture.

A further consideration is that continuous transition through a set of sectional images can be considered to be a form of animation. Several studies have suggested that computer animation can benefit learning, particularly when the user has control of the presentation (Kaiser, Proffitt, Whelan, & Hecht, 1992; Pani, Chariker, Dawson, & Johnson, 2005, Experiment 2; Schwartz, Blair, Biswas, Leelawong, & Davis, 2007). On the other hand, other studies have shown that animation is not necessarily advantageous for learning (Mayer, Hegarty, Mayer, & Campbell, 2005; Pani, Chariker, Dawson, & Johnson, 2005, Experiment 1; Tversky, Morrison, & Betrancourt, 2002). Overall, it seemed an open question whether a capability for continuous navigation through sets of anatomical sections would benefit learning.

Design Decisions for Computer-Based Instruction

Ecological Validity

A central goal for this project was to design research that would generalize to real learning situations (see McNamara, 2006; Barab, 2006). In particular, we covered a curriculum in neuroanatomy that represented a scale of difficulty one might encounter in the early parts of an introductory course in neuroscience. Materials and procedures were sufficiently realistic and comprehensive that they could be deployed in actual neuroanatomy instruction. The participants experienced a situation much like what beginning students would encounter. Thus, they approached the topic as novices, and they were trained to a high performance criterion. Learning occurred over a period of weeks and required regular visits to the lab.

Adaptive Exploration

The learning programs implemented an approach to user interaction in computer-based learning that we call *adaptive exploration*. In adaptive exploration, tools are provided for flexibly exploring the content of a domain. For example, a learner might have the capability to rotate an anatomical model and to remove structures so as to see behind them. In addition, these tools are presented in the context of repeated trials of study, test, and feedback. As a consequence, the participant learns the nature of the test of performance that must be mastered and the degree of progress toward achieving satisfactory performance. Over time, exploration is used to provide the information required for successful learning in that domain. Note that the concepts of adaptation (Graesser, Jackson, & McDaniel, 2007; Koedinger & Corbett, 2006) and guided exploration (Klahr & Nigam, 2004; van der Meij & de Jong, 2004) are well represented in previous work. Our development of adaptive exploration is an effort to apply these fundamental concepts in a complex spatial domain.

The version of adaptive exploration implemented for this study serves two additional purposes. First, it conforms to what appear to be best practices for optimizing long-term retention through spaced practice (Melton, 1970), repeated testing (Karpicke & Roediger, 2008), and feedback (Kornell, Hays, & Bjork, 2009). Across multiple trials distributed over time, participants can adaptively focus exploration during study, but they are always tested and provided feedback on the entire set of items to be learned.

Second, we were concerned that neuroanatomy includes numerous complex three-dimensional shapes and configurations. We thought that adaptive exploration would promote subjective organization (Tulving, 1962), discovery of informative views of anatomy, and adaptive learning of complex configurations. We acknowledge, however, that it is sometimes better to provide learners with a few key views of three-dimensional structures rather than to encourage full rotational exploration (Garg, Norman, Eva, Spero, & Sharan, 2002; Levinson, Weaver, Garside, McGinn, & Norman, 2007). Moreover, full user interaction with material is not always the optimal method for learning (Keehner, Hegarty, Cohen, Khooshabeh, & Montello, 2008).

Evaluation of Learning

In the comparison of methods of instruction, a common concern is whether time or effort with the material to be learned has been equated (Tversky et al., 2002). For the current study, there were two primary choices with regard to experimental control. Time with the material could be held constant across the experimental groups, or the level of performance reached during learning could be held constant. Constraining time meant that participants were likely to reach different levels of performance at the end of learning. Constraining the level of performance meant that participants were likely to take different amounts of time to learn. Three considerations suggested that a single high level of performance should be enforced before learning was stopped and that time to learn should be permitted to vary. First, the primary hypothesis of transfer of learning from whole to sectional anatomy was focused on the efficiency of initial learning. Thus, the time required to learn the material was best measured as an outcome rather than constrained to be constant. Second, long-term retention and the transfer of knowledge to the interpretation of biomedical images were intended as measures of external validity. For these tests to be meaningful, initial learning had to be successful and uniform across conditions. Third, in advanced disciplines such as neuroanatomy, the typical student will study the material until it is known well. Thus, considerations of ecological validity suggested a uniformly high performance criterion.

A potential criticism of this design decision is that differences in the amounts of experience with material might ultimately be related to differences in long-term retention and transfer of knowledge. This can be considered to be a confounding variable for any other differences between experimental conditions. One part of the response to this criticism is that several studies indicate that additional study time by itself does little to improve retention of material (Karpicke & Roediger, 2008; Nelson & Leonesio, 1988: Zimmerman, 1975). On the other hand, the literature on the "test effect" indicates that long-term retention is related to the number of tests during learning (Johnson & Mayer, 2009; Karpicke & Roediger, 2008; Nungester & Duchastel, 1982; Roediger & Karpicke, 2006).

The implications of this extensive body of research for the present study are that the retention of knowledge of sectional anatomy would be related to the number of tests of sectional anatomy during learning. Because it was hypothesized that a transfer of learning from whole to sectional anatomy would lead to fewer trials of sectional anatomy without a cost for long-term retention or transfer of knowledge, the potential confound with the

test effect was considered to be an acceptable risk. If, on the other hand, additional experience with whole anatomy led to better long-term retention of sectional anatomy, this would be a further instance of the hypothesized transfer of learning from whole to sectional anatomy (see Rohrer, Taylor, & Sholer, 2010). This issue will be reconsidered in the Discussion in the context of the experimental data.

Method

Participants

Seventy-two undergraduate students at the University of Louisville were recruited for the study (31 male and 41 female). All were at least 18 years of age. Three of the 72 participants dropped out of the study after completing learning but before completing all testing. Participants were recruited through advertisements placed around the university campus and on university websites. On their first visit to the study, volunteers were surveyed about their knowledge of the 19 neuroanatomical structures taught in the study. Only those respondents who reported minimal knowledge were enrolled as participants. Participants were paid $8 per hour for their participation.

Psychometric Tests

Prior to the experiment, each participant was administered the Space Relations subtest of the Differential Aptitude Tests, a standard test of spatial ability (DAT-SR; Bennett, Seashore, & Wesman, 1989). Differences in spatial ability were a potential source of variability in the data (Keehner et al., 2008; Pani, Chariker, Dawson, Johnson, 2005). Therefore, the mean and the distribution of spatial ability scores were balanced across the four learning groups (i.e., a stratified design). Participant spatial ability scores were normally distributed and ranged from the 5th to the 99th percentile, with a mean at the 61st percentile. Differences in spatial ability among the four groups were small and not statistically significant.

Materials

Neuroanatomical model. A three-dimensional computer graphical model of the human brain, shown in Figure 1, was created for this research. The model was created to be relatively complete and sufficiently accurate when sectioned. Accuracy was a major concern due to the planned tests of transfer to the interpretation of biomedical images. Digital photographs of cryosections of the human head in the Visible Human project (Version 2.0) of the National Library of Medicine were used as source material to create the model (Ackerman, 1995; Ratiu, Hillen, Glaser, & Jenkins, 2003).

The brain model was composed of the cerebral cortex with well defined gyri and sulci, ventricles, cerebellum, brainstem with associated white matter, and 15 additional subcortical structures. These included the amygdala, caudate nucleus, fornix, globus pallidus, hypothalamus, hippocampus, mammillary bodies, nu-
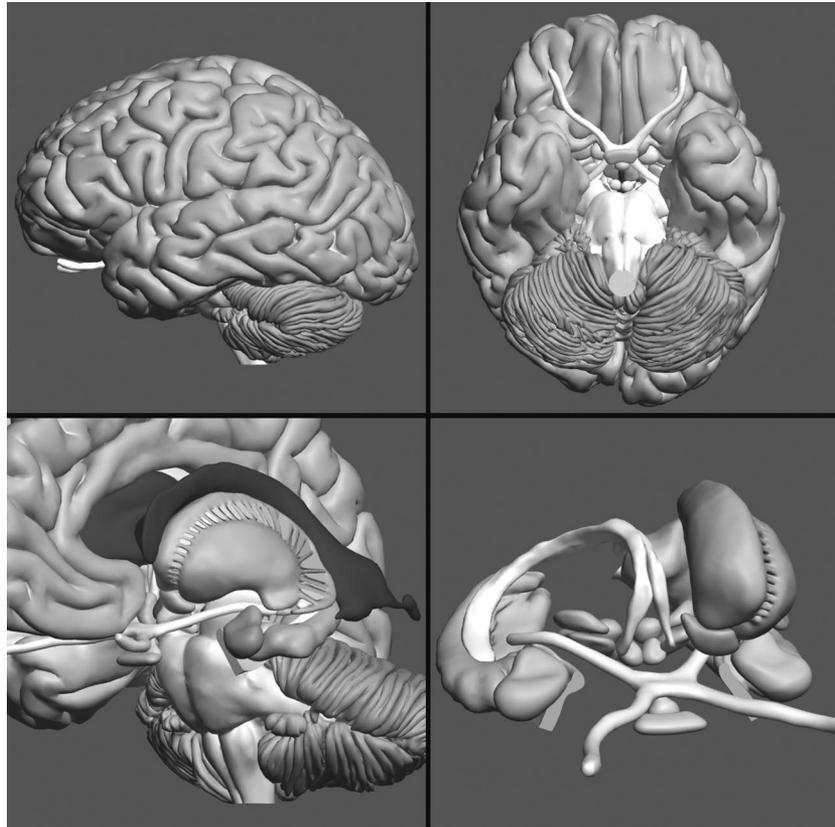

Figure 1. Four views of the computer graphical model of neuroanatomy created for this project.

cleus accumbens, optic tract, pituitary, putamen, red nucleus, substantia nigra, subthalamic nucleus, and thalamus. The structures were colored in dark gray (ventricles), medium gray, and white to approximate the basic appearance of light and dark structures in typical biomedical images of the brain.

Neuroanatomical sections. Three relatively dense sets of serial sections were created from the brain model. There were 60 coronal sections (i.e., in a "front" view), 50 sagittal sections ("side" view), and 46 axial sections ("top" view), with all sections taken at equal intervals.

The MRI images used to test transfer were made available from the SPL-PNL Brain Atlas (Kikinis et al., 1996). The images are typical low-resolution gray scale T1 images of structures in the head and neck. They contain many structures not in the brain, such as muscle and bone. The images in the data set are densely sampled with a resolution of 256×256 pixels in all three viewing planes. The images were slightly brightened, contrast enhanced, and presented at a screen resolution of 895×895 pixels.

The Visible Human images used to test transfer were from the Visible Human 2.0 data set used to create the brain model. The images are high-resolution color photographs of cryosections of the head and neck. They clearly depict many structures not in the brain, such as muscle, bone, connective tissue, and blood vessels. The images were presented at screen resolutions of 657×919 (coronal), 972×919 (sagittal), and 672×953 pixels (axial).

Learning programs. Computer programs for learning neuroanatomy were created using the C++ programming language and the Open Inventor library for interactive graphics. Participants interacted with virtual neuroanatomy using a standard mouse to select structures and to operate interactive tools that were displayed on the screen. There was a common format and interface for all of the learning programs. The differences between the programs were modifications related to the presentation of anatomy. The common format will be described first.

A participant completed two learning trials, one block of trials, before a single run of the program terminated. Participants were presented with the same view of anatomy (e.g., coronal) throughout the two trials in a block. Each learning trial was composed of three phases: study, test, and feedback. Screenshots from the study phase in whole and sectional anatomy learning are presented in Figures 2 and 3.

In the study phase of a learning trial, participants had 3 min to freely explore the brain using tools that functioned specifically for either whole or sectional anatomy. Participants could learn the names of neuroanatomical structures by selecting them. Upon selection, a structure was highlighted in yellow, and its name was displayed prominently below the model. A panel on the right side of the screen displayed labeled buttons for each of the anatomical structures in the model and included separate buttons for left and right structure pairs.

In the test phase of a learning trial, the participant's task was to identify the 19 structures in the model. Participants selected a structure and then selected the appropriately labeled button from the button panel. Only primary structure names appeared on the buttons (i.e., left and right were not indicated). Once a name was submitted, the color of the anatomical structure changed to blue to indicate a completed answer. Participants could choose to skip structures that they could not identify. Testing was self-paced. After identifying as many of the 19 structures as possible, participants exited the test phase.

Immediately after the test phase was completed, a message panel at the center of the screen provided numerical feedback regarding the number of structures named correctly, the number named incorrectly, and the number omitted. Participants then
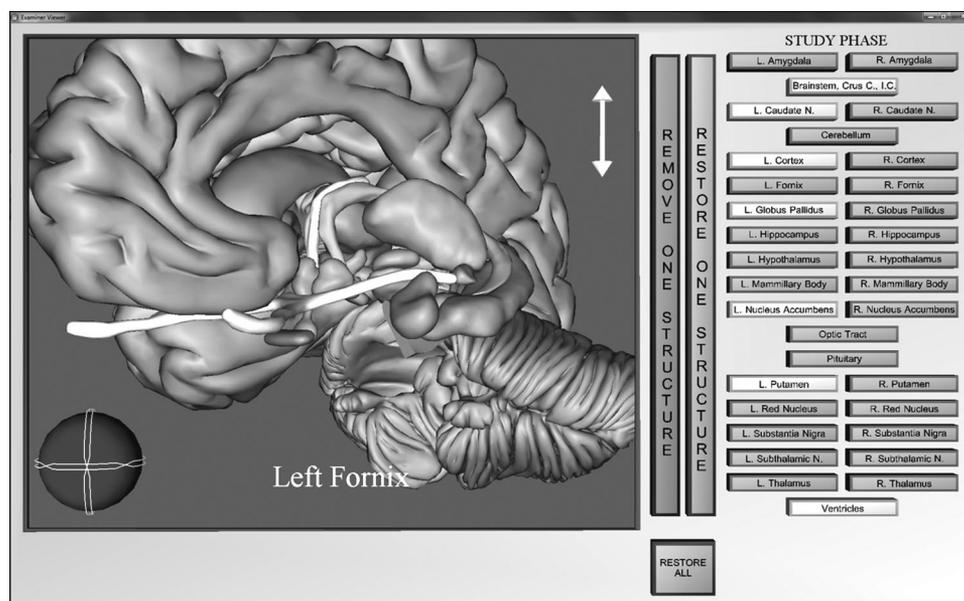
Figure 2. Screenshot of the anatomical model and the interface in the study phase of the whole anatomy learning program. The learner has removed several structures, rotated the model, and zoomed in to view the left fornix. Note that gray scale reproductions of this figure do not preserve the visual salience of selected structures.
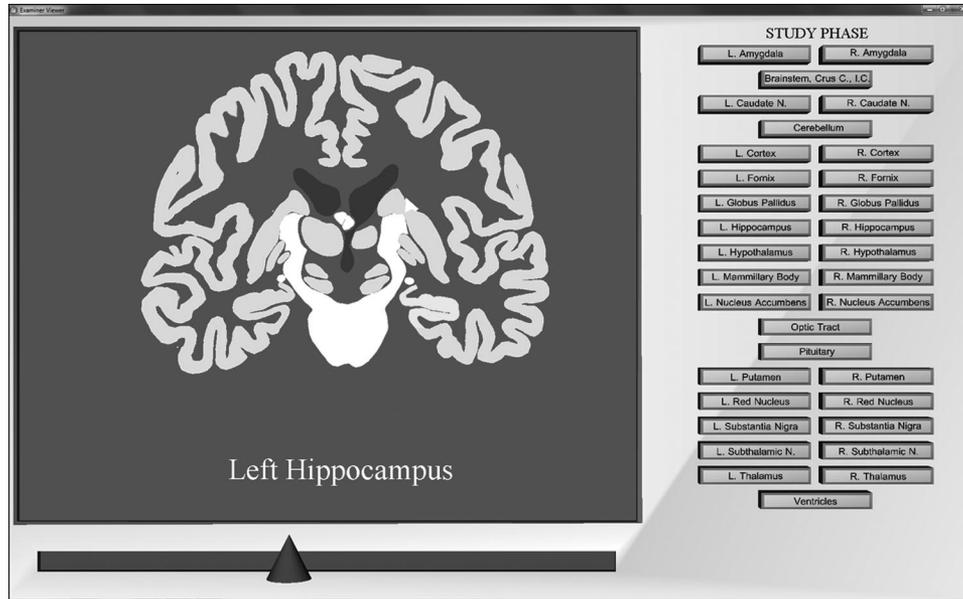
Figure 3. Screenshot of the anatomical model and the interface in the study phases of the sectional anatomy learning programs. The learner has moved the slider to reveal a middle section in the coronal view and has selected the left hippocampus. Note that gray scale reproductions of this figure do not preserve the visual salience of selected structures.

proceeded to an interactive feedback phase of the trial (simply called feedback phase).

In the feedback phase, the participant saw the same orientation of the brain and used the same tools and procedures as in the study phase. In the feedback phase, however, structures were color coded to provide participants with information about their performance on the test. Structures that had been named correctly were colored green. Structures that had been named incorrectly were colored red. Structures that had been omitted remained their original colors. Participants had 3 min to explore the brain in the context of the graphical feedback.

Program for whole anatomy learning. In the study and feedback phases for whole anatomy learning, a rotation tool allowed participants to smoothly rotate the model 360° forward and backward or right and left. A zoom tool allowed participants to move the model closer or further from view. An easily accessible button permitted participants to remove a selected structure. While a structure was gone, the labeled button was white and appeared to be pushed in. Buttons were also available for restoring structures, individually or together.

In the test phase of a trial, model rotation was constrained to a total range of 90° of motion, 45° in any direction from the initial viewpoint. This ensured that a participant's performance on the test was specific to the viewpoint being learned in that trial. Removing and restoring structures was necessary so that participants could find and label them.

Programs for sectional anatomy learning. Two programs were created for learning sectional anatomy, one for the continuous and one for the discrete form of navigation. In the study phase of a trial, both programs presented a set of anatomical sections in serial order in a single viewing plane. As usual, participants could select a structure, highlight it, and learn its name. There was a

slider at the bottom of the screen that was used to navigate through the sections. The two learning programs differed in how navigation through the sections took place and in whether the highlighting of structures remained during navigation.

In the continuous program, moving the slider resulted in continuous movement back and forth through the series of sections. A section of the brain was always visible, and the transition between sections comprised a type of animation. In the program for discrete presentation, movement between sections was perceptually discontinuous. When participants moved the slider, the brain became invisible. The standard background could be seen behind the interface, and the number of the corresponding section in the series appeared prominently at the bottom of the viewing area. On stopping at a particular numbered section, a 0.75 s delay occurred before the appropriate section of the brain appeared on the screen.

As the slider was moved in the continuous program, a highlighted structure remained highlighted in each section in which it appeared. In the discrete program, when participants moved to a new section, highlighting of structures was removed.

The test phase of a learning trial was the same in the two sectional learning programs. Participants were given a series of test sections presented one at a time. In each section, one or more structures were indicated with a red arrow, and the participant's task was to correctly label the structures under the arrows.

During the feedback phase of the trial, participants used the slider to find each of the test sections in the series. A message reading "Test Section" appeared prominently on the screen when a test section was accessed by the slider. The tested structures in each test section were identified with the same red arrows that appeared in the test.

Criteria for selection of test structures. The goal for the tests was that each one should provide an accurate assessment of

a participant's knowledge of the 19 neuroanatomical structures. Three criteria were used to create a set of sectional anatomy tests. First, each of the 19 structures was sampled at least once in each test. Second, structures that spanned more than seven sections were sampled twice, once from the first half of the series and once from the second half. This meant that for the coronal and sagittal views, 10 structures were tested twice, and nine were tested once (29 total). For the axial view, eight structures were tested twice, and 11 were tested once (27 total). The third criterion required that as much as possible, different sections from individual structures were sampled across the tests. Eight unique sectional anatomy tests were created for each of the three views of anatomy.

Test scoring. The learning programs calculated percentage correct for individual test trials and mean percentage correct for the two trials in each block. To equate scoring across whole and sectional anatomy tests and across structures of different sizes within sectional anatomy, single test items were defined at the level of single neuroanatomical structures. Thus, if a single anatomical structure was tested with two sectional views, each view was given a weight of 0.5 structures. Although the intent was to create an unbiased scoring system, any possible bias would appear to favor sectional tests and the sections alone groups.

Transfer tests: MRI and Visible Human images. Three computer programs were developed to test transfer of knowledge to the interpretation of biomedical images. In the first test, Uncued Recognition, nine biomedical images of the same type, MRI or Visible Human, were presented one at a time. For a single image type, the images alternated between coronal, sagittal, and axial views. Across the nine images, a sectional image was presented from the early, middle, and latter portions of the brain for each of the views. For each structure that was recognized in an image, the participant indicated the location with the mouse (leaving a red dot on the image) and labeled the structure with the button panel. After the test was completed for one type of image, the participant proceeded to the other type.

The remaining two test programs, Submit Structure and Submit Name, provided cues to the presence of structures in the images. In the Submit Structure test, the name of a single structure was presented at the bottom of each image, and participants clicked on the appropriate structure in the image. In the Submit Name test, a single structure was designated by a red arrow in each image, and participants selected the structure name from a button panel.

The Submit Name and Submit Structure tests were each administered in six subtests. A subtest presented either MRI or Visible Human images in a coronal, sagittal, or axial view. MRI and Visible Human images were paired for each view, with the order of image type counterbalanced. The subtests were created using the same selection criteria as were used to create tests in the sectional anatomy learning programs. Thus, the number of test items in the coronal and sagittal views was 29, and the number of test items in the axial view was 27.

Retention or transfer tests. Tests were administered to measure the long-term retention of sectional anatomy and whole anatomy. For participants who had only seen sectional anatomy, the test of whole anatomy was a test of transfer rather than retention. These tests were the same as those given during the learning trials for sectional and whole anatomy, but without study and feedback. All three views were tested for both representations of anatomy.

Instruction programs. A set of instructional programs was developed to train participants in the use of the software prior to beginning the study. To ensure that learning neuroanatomy did not begin during these demonstrations, a "mock brain" was used for this instruction. Although it was fictitious, it reflected major structural properties of a real brain, such as bilateral symmetry, variation of size and shape, and placement of some structures inside of others. It was presented whole or in sections, depending on need.

Apparatus

Participants sat individually at computer workstations with large high resolution liquid crystal display screens (24 in. [60.96 cm], $1,200 \times 1,952$ pixels). The workstations had high capacity graphics capability that allowed smooth high quality presentation of three-dimensional graphics. Participants were tested alone in small quiet rooms with the doors closed and the room lights on.

Design and Procedure

Experimental design. The core experimental design was a 2×2 between-groups factorial. One factor, anatomy course (AC), varied whether participants learned only sectional anatomy (the sections alone groups) or learned whole anatomy followed by sectional anatomy (the whole plus sections groups). The second factor, sectional anatomy presentation (or SA presentation), varied whether exploration of sectional anatomy was continuous or discrete.

During the learning part of the study, performance in identifying 19 neuroanatomical structures was measured over multiple blocks of trials for either whole anatomy or sectional anatomy. Because performance related to view was not a central question for this study, the order in which views were presented over blocks of trials was not counterbalanced across participants. Rather, it was decided to minimize unnecessary variability in performance by standardizing the order among the views at coronal, followed by sagittal, and then axial. This alternating order was intended to provide the benefits for learning that arise from interleaving different types of material (Kornell & Bjork, 2008; Rohrer & Taylor, 2007).

Percentage correct was calculated for each trial during learning, and mean percentage correct was calculated for each block of two trials. Participants continued learning anatomy in a single presentation (whole or sectional anatomy) until they reached a minimum of 89.5% accuracy in each of three consecutive learning blocks—all three views of anatomy. This value, 89.5%, corresponded to 17 of the 19 structures to be learned.

Immediately after learning was completed, participants were given the three tests of transfer to biomedical images in the order Uncued Recognition, Submit Structure, and Submit Name. The order was intentionally not counterbalanced so that the tests that were likely to be easier and more informative would not contaminate the tests that were likely to be more challenging. Participants were instructed prior to taking these tests that the MRI and Visible Human images would vary from the model they were shown in numerous ways and that they should use their knowledge of neuroanatomy to make the best judgment they could about the structures in the images. No feedback on performance was given for these tests.

For each of these tests, the order of testing with the two image types, MRI and Visible Human, was counterbalanced across participants. For each type, participants were tested on all three views of anatomy in the standard order: coronal, sagittal, and axial. This resulted in a $2 \times 2 \times 2 \times 3$ design (AC $\times$ SA Presentation $\times$ Image Type $\times$ View/Order) for each test.

Two to 3 weeks after learning was completed, participants were given the test of long-term retention for sectional anatomy followed by the test of long-term retention or transfer for whole anatomy. For each, participants were tested with coronal, sagittal, and axial views, in that order. This resulted in a $2 \times 2 \times 3$ design (AC $\times$ SA Presentation $\times$ View/Order) for each test.

Chronology for individual participants. On the first visit to the lab, participants completed a questionnaire that surveyed their degree of familiarity with each of the 19 neuroanatomical structures. Participants then completed the DAT-SR and were placed in an experimental group so as to stratify spatial ability. Following placement, each participant was given a short orientation to the global anatomy of the brain and the three sectional views of the brain, using line drawings. Afterward, the entire course of learning and testing was explained to the participants. Participants were then given specific instructions and practice using the mock brain for the learning task that they would use to begin instruction in neuroanatomy.

On the next visit to the lab, participants went through the instruction programs a second time and were required to demonstrate understanding of the procedures and the software tools. Then they began the learning portion of the study. After participants in the whole plus sections groups reached the learning criterion for whole anatomy, they were given instructions for using the sectional anatomy programs and then began those trials. Participants visited the lab a minimum of two times per week for an hour each time, with an average rate of three visits per week. After learning was completed, participants continued on their regular schedules to complete the tests of transfer of knowledge to interpreting biomedical images. Participants returned to the lab 2–3 weeks later to take the tests of long-term retention or transfer. On average, participants completed the study in 5 weeks (SD = 1.4 weeks, range = 2.9–9.1 weeks).

Results

Multilevel modeling was used for statistical analysis of performance in learning (Raudenbush & Bryk, 2002; Singer & Willett, 2003). One advantage of this approach is its capacity for handling unbalanced data. Participants reached performance criteria at different points in time, and multilevel modeling could be used to analyze these data appropriately. Binomial multilevel models were most appropriate for these data (Rice, 2001), and effect sizes are expressed in odds ratios (ORs) derived from log odds coefficients. Details of model parameters are available from the authors.

Learning

Relative efficiency of learning whole and sectional anatomy. The value of transfer of learning from whole to sectional anatomy will depend in part on whether learning whole anatomy is efficient. Therefore, a basic question for this research was whether learning whole anatomy was more efficient than learning sectional anatomy

without the benefit of transfer. To answer this question, performance in learning whole anatomy (in the whole plus sections groups) was compared with performance in learning sectional anatomy without transfer (the sections alone groups). Variables tested for inclusion in the multilevel model included trial block, anatomy presentation (whole or sectional anatomy), and DAT-SR (spatial ability). Mean percentage correct identification across blocks of trials in whole anatomy learning and sectional anatomy learning (without transfer) is presented in Figure 4 for all four groups.

Participants who started the experiment by learning whole anatomy had substantially higher performance in the first block of trials than participants who started by learning sectional anatomy, $t(69) = 5.780$, $p < .001$, $OR = 1.985$. Both groups improved in performance over successive trial blocks; however, the increase in performance was much greater for participants learning whole anatomy: block, $t(68) = 15.746$, $p < .001$, $OR = 1.582$; anatomy presentation, $t(68) = 7.359$, $p < .001$, $OR = 2.476$.

An analysis of variance was performed to compare the number of blocks of trials required to complete learning in whole anatomy (from whole plus sections) and sectional anatomy (sections alone). Spatial ability was correlated with the number of blocks required ($r = -.236$, $p = .046$) and was entered as a covariate, $F(1, 67) = 7.785$, $p = .007$, $\eta_p^2 = .104$. Participants learned whole anatomy in significantly fewer blocks ($M = 5.2$, $SD = 3.1$) than participants learned sectional anatomy ($M = 10.7$, $SD = 3.1$), $F(1, 67) = 57.555$, $p < .001$, $\eta_p^2 = .462$.

Efficiency of learning sectional anatomy with continuous or discrete control. There were no effects on the efficiency of learning sectional anatomy due to sectional anatomy presentation (i.e., continuous or discrete navigation through the sections) in any of the analyses of learning. This variable was not retained in the
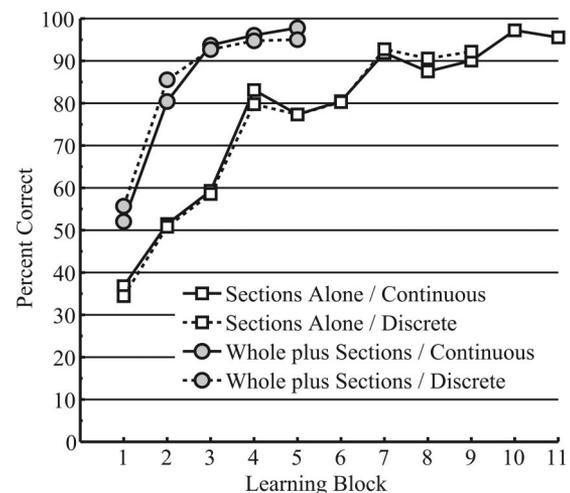

Figure 4. Mean percentage correct test performance during learning for the initial presentation of whole and sectional anatomy, broken down by learning block and learning group. Means for whole anatomy come from the whole plus sections groups. Means for sectional anatomy come from the sections alone groups.

multilevel models, and it will not be discussed further in the presentation of results on learning.

Transfer of learning from whole to sectional anatomy. Transfer of learning from whole to sectional anatomy was first explored with a comparison of performance in sectional anatomy learning between the whole plus sections and sections alone groups. If there was transfer, the whole plus sections groups should have performed better than the sections alone groups in the first block of sectional learning. Mean percentage correct identification across blocks of trials in sectional anatomy learning is presented in Figure 5 for all four groups.

Participants in the whole plus sections groups performed substantially better in the first block of sectional anatomy learning than participants in the sections alone groups (see Figure 5), $t(69) = 13.522$, $p < .001$, $OR = 4.173$. Although both groups improved over time, the whole plus sections groups continued their learning at a slower rate than the sections alone groups: block, $t(70) = 16.021$, $p < .001$, $OR = 1.543$; AC, $t(70) = -3.321$, $p = .002$, $OR = 0.889$.

An analysis of variance was performed to compare the whole plus sections groups and the sections alone groups on the number of trial blocks required to reach criterion in sectional anatomy learning. Spatial ability was correlated with the number of blocks to reach criterion ($r = -.298$, $p = .011$) and was included as a covariate, $F(1, 67) = 7.678$, $p = .007$, $\eta_p^2 = .103$. On average, participants in the whole plus sections groups completed sectional anatomy in 2.5 fewer blocks ($M = 8.2$, $SD = 3.9$, range: 4–19) than participants in the sections alone groups ($M = 10.7$, $SD = 3.9$, range: 6–26), $F(1, 67) = 7.282$, $p = .009$, $\eta_p^2 = .098$.

The time required to complete the test within each trial was under the control of the participants. Across all groups, participants spent an average of 4 min 48 s per test in sectional anatomy learning, 10.15 s per test item. It is interesting that participants in the whole plus sections groups took somewhat longer to test themselves (whole plus sections, $M = 5$ min 6 s, $SD = 1.1$ min; sections alone, $M = 4$ min 29 s, $SD = 0.9$ min; difference = 36 s or 1.28 s per test item), $t(70) = 2.467$, $p = .016$, Cohen's $d = 0.59$.
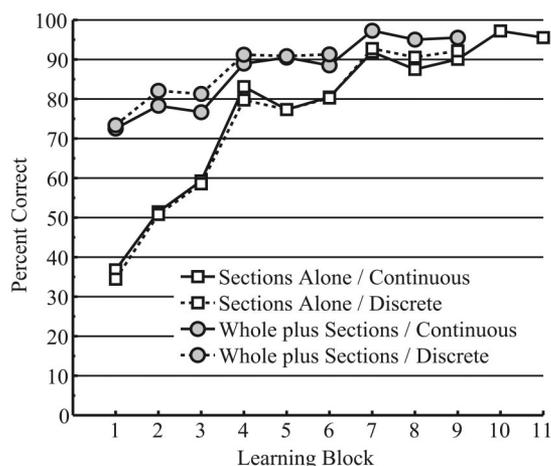

Figure 5. Mean percentage correct test performance in sectional anatomy learning, broken down by learning block and learning group. Participants in the whole plus sections groups completed learning of whole anatomy prior to learning sectional anatomy.

Clearly, the high rate of transfer required effort on the part of the whole plus sections participants. On the other hand, because the whole plus sections groups required fewer blocks of trials to complete sectional anatomy learning, their total test time across all of sectional anatomy learning was actually somewhat less than the sections alone groups, although statistically they were not different (whole plus sections, $M = 85$ min 3 s, $SD = 47.3$ min; sections alone, $M = 96$ min 43 s, $SD = 47.3$ min; $p = .294$).

Comparison of whole plus sections and sections alone groups for total neuroanatomy learning. Differences between conditions were further explored by comparing performance in sectional anatomy for the whole plus sections and sections alone groups after relating performance to the total time spent learning neuroanatomy. For the whole plus sections groups, learning blocks were numbered to reflect the time participants spent learning both whole and sectional anatomy. Nearly two thirds of the participants in the whole plus sections group (21 of 36) completed whole anatomy learning in four blocks and transferred to sectional anatomy in Block 5. No one completed whole anatomy in fewer than four blocks. Therefore, multilevel modeling was applied to performance in sectional anatomy learning beginning at Block 5. Modeled performance at Block 5 was 71% correct identification for the whole plus sections groups and 80% for the sections alone groups, $t(69) = -3.030$, $p = .004$, $OR = 0.615$. The 9% difference is equivalent to two of the 19 structures on the test.

An analysis of variance was performed to look for differences between the groups in the number of blocks of trials necessary to complete all learning in neuroanatomy. Spatial ability was correlated with the number of blocks to reach criterion ($r = -.344$, $p = .003$) and was included as a covariate, $F(1, 67) = 10.129$, $p = .002$, $\eta_p^2 = .131$. Participants in the whole plus sections groups completed whole anatomy and sectional anatomy in 2.7 more blocks than participants in the sections alone groups completed sectional anatomy (whole plus sections, $M = 13.4$, $SD = 4.6$; sections alone, $M = 10.7$, $SD = 4.6$), $F(1, 67) = 6.021$, $p = .017$, $\eta_p^2 = .082$. By way of comparison, the mean time to complete whole anatomy learning was 5.2 blocks, with a modal completion time of four blocks.

Total error in learning neuroanatomy. An alternative to counting the number of trials for gauging the efficiency of learning is to count the errors committed by participants during testing. Test errors, including omissions, were summed across all learning of neuroanatomy for each participant. Participants in the whole plus sections groups committed significantly fewer errors than participants in the sections alone groups (whole plus sections total error: $M = 77$, $SD = 49.5$; sections alone total error: $M = 100$, $SD = 49.5$), $F(1, 67) = 3.870$, $p = .053$, $\eta_p^2 = .055$. The smaller level of total error for the whole plus sections groups is especially remarkable considering that these groups took 2.7 more blocks to complete learning. These groups met the performance criteria twice and took somewhat longer to do so, but spent a greater proportion of the time at high levels of performance. Spatial ability was a significant covariate in the analysis of total error, $F(1, 67) = 13.995$, $p < .001$, $\eta_p^2 = .173$.

Long-Term Retention

Multivariate analysis of variance with Bonferroni correction was used to analyze the data for the tests of long-term retention or

transfer and for the tests of transfer to the interpretation of bio-medical images.

Long-term retention of sectional anatomy.   Retention of sectional anatomy remained high 2–3 weeks after learning, with several participants reaching 100% accuracy in the first test. Mean percentage correct identification was 88.5% for the whole plus sections groups and 86.5% for the sections alone groups. There was a main effect of view/order, Wilks's $\Lambda = .819$, $F(2, 63) = 6.949$, $p = .002$, $\eta_p^2 = .181$. Participants performed better with the initial coronal view ($M = 91.5\%$, $SD = 1.0$), compared with the later sagittal ($M = 85.4\%$, $SD = 0.9$), $t(68) = 5.491$, $p = .003$, $d = 0.66$, and axial views ($M = 85.5\%$, $SD = 1.0$), $t(68) = 5.158$, $p = .003$, $d = 0.62$.

There was an interaction of AC with view, $\Lambda = .898$, $F(2, 63) = 3.570$, $p = .034$, $\eta_p^2 = .102$. The whole plus sections groups retained 4.7% more than the sections alone groups for the sagittal view of sectional anatomy (whole plus sections: $M = 87.8\%$, $SD = 1.3$; sections alone: $M = 83.1\%$, $SD = 1.3$), $t(57) = -2.675$, $p = .03$, $d = 0.71$. There were no differences between the groups in retention of the coronal and axial views.

Long-term retention and transfer for whole anatomy.   Participants in the whole plus sections groups were more accurate than participants in the sections alone groups in identifying whole anatomy, $F(1, 64) = 15.306$, $p < .001$, $\eta_p^2 = .193$. Indeed, participants in the whole plus sections groups tested at 97% mean accuracy ($SD = 1.4$) in identifying whole brain structures. There was again a main effect of view/order on percentage correct identification (coronal: $M = 92\%$, $SD = 1.3$; sagittal: $M = 93.3\%$, $SD = 1.2$; axial: $M = 94.9\%$, $SD = 1.0$), $\Lambda = .904$, $F(2, 63) = 3.335$, $p = .042$, $\eta_p^2 = .096$.

Although participants in the sections alone groups had never seen whole anatomy, they reached an overall mean accuracy of 89.5% ($SD = 1.4$). This meets the numerical criterion used for successful learning. On the other hand, a relatively substantial effort was required to achieve this performance. All tests in this experiment were self-paced. In an analysis of test duration, participants in the sections alone groups took nearly 2 min longer than the whole plus sections groups to complete each test for whole anatomy (whole plus sections: $M = 3.20$ min, $SD = 0.21$; sections alone: $M = 5.16$ min, $SD = 0.22$), $F(1, 61) = 54.331$, $p < .001$, $\eta_p^2 = .471$. This suggests that participants who had learned only sectional anatomy were not recalling a representation of whole anatomy but rather were able to infer the identities of whole structures from what they had learned about sectional anatomy.

Transfer of Learning to the Interpretation of Biomedical Images

Scoring.   Two of the tests of transfer of learning to the interpretation of biomedical images, Uncued Recognition and Submit Structure, required participants to indicate the locations of structures on the images. To score these responses, areas on the images corresponding to correct answers were decided ahead of time, and images were created with the structure boundaries drawn on them. The participants' answers were then superimposed onto the new images and scored against the explicit boundaries. When scoring the responses, the experimenters were blind to the participants' identities and experimental conditions.

For Uncued Recognition, the number of structures identified correctly out of the total number possible was tabulated for each image, and percentage correct over each combination of image type and view (e.g., coronal MRI) was recorded for each participant. In Submit Structure and Submit Name, the test design was similar to the tests of sectional anatomy learning in the learning stage of the study, and responses were tabulated and scored in the same manner.

Analysis.   There were no differences related to AC (whole plus sections vs. sections alone) in the transfer of knowledge to the interpretation of biomedical images. On the other hand, in two of the three tests of transfer, Uncued Recognition and Submit Name, there was a main effect of SA presentation (continuous vs. discrete presentation of sectional anatomy): Uncued Recognition, $F(1, 65) = 3.962$, $p = .051$, $\eta_p^2 = .057$; Submit Name, $F(1, 64) = 4.835$, $p = .032$, $\eta_p^2 = .070$. It is interesting that participants who learned sectional anatomy using a discrete presentation were slightly more accurate in identifying structures than participants who learned with a continuous presentation. Mean percentage correct for the interpretation of biomedical images, broken down by test type, image type, sectional anatomy presentation, and view/order is presented in Figure 6.

Performance varied widely among individuals, with some participants performing extremely well. In Submit Structure and Submit Name, the two tests that provided cues, several of the best performing participants were above 90% accuracy. In all three tests, performance was substantially higher for the Visible Human images than the MRI images, suggesting that the quality of the imagery is important to the level of transfer: Uncued Recognition, $\Lambda = .440$, $F(1, 65) = 82.659$, $p < .001$, $\eta_p^2 = .560$; Submit Structure, $\Lambda = .704$, $F(1, 64) = 26.913$, $p < .001$, $\eta_p^2 = .296$; Submit Name, $\Lambda = .378$, $F(1, 64) = 105.282$, $p < .001$, $\eta_p^2 = .622$. Given that these are tests of the transfer of learning, mean performance for the Visible Human images in the Submit Structure and Submit Name tests was quite impressive (see Figure 6).

There was also a main effect of view/order in all three tests: Uncued Recognition, $\Lambda = .868$, $F(1, 65) = 4.853$, $p = .011$, $\eta_p^2 = .132$; Submit Structure, $\Lambda = .884$, $F(2, 63) = 4.118$, $p = .021$, $\eta_p^2 =$
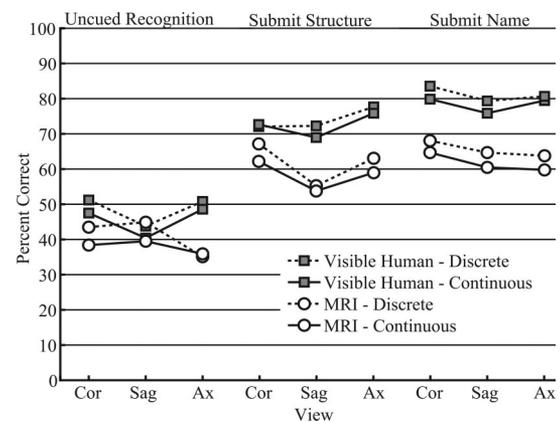

Figure 6.   Transfer of learning to the interpretation of biomedical images: Mean percentage correct test performance for test type, image type, view, and sectional anatomy presentation. Cor = coronal; Sag = sagittal; Ax = axial; MRI = magnetic resonance imaging.

.116; Submit Name, $\Lambda = .849$, $F(2, 63) = 5.597$, $p = .006$, $\eta_p^2 = .151$. Once again, the sagittal view of anatomy appeared to be the most consistently challenging: Uncued Recognition, coronal $>$ sagittal, $t(68) = 3.072$, $p = .009$, $d = 0.37$; Submit Structure, coronal $>$ sagittal, $t(68) = 4.915$, $p = .003$, $d = 0.59$; Submit Structure, axial $>$ sagittal, $t(68) = -5.344$, $p = .003$, $d = 0.64$; Submit Name, coronal $>$ sagittal, $t(68) = 3.699$, $p = .003$, $d = 0.45$.

Discussion

Neuroanatomy is an advanced discipline focused on complex spatial structures. It is a difficult discipline to master, particularly with the need to learn sectional representation. Against this backdrop, it can be said that all groups in this study learned neuroanatomy efficiently. Consider, for example, that a single learning trial included 3 min of study, a self-paced test, and 3 min of feedback. In about 25 such trials, novices learned consistently to recognize 19 neuroanatomical structures in whole form and in the three standard sectional views. This outcome is encouraging for the development of computer-based learning using the methods of adaptive exploration.

The experimental results provided clear evidence that instruction of neuroanatomy should be designed on the basis of substantial transfer of learning from whole to sectional anatomy. Whole anatomy is learned in half the time that it takes to learn the corresponding sectional anatomy. Once learned, knowledge of whole anatomy transfers well to learning sectional anatomy. In particular, performance in the early trials of sectional learning is substantially higher than it would have been, and instruction of sectional anatomy is completed faster. Learning two forms of anatomical representation does take slightly longer than learning one, but there is less error over the total course of learning. There was no disadvantage of using this instructional method for long-term retention of sectional anatomy or for transfer of knowledge to the interpretation of biomedical images.

It is unlikely that the transfer of learning from whole to sectional anatomy is due to learners imagining the outcome of sectioning whole anatomy. The literature on reasoning about spatial transformations suggests that this is a mental process that most people could not apply accurately to anatomical relations (Hegarty & Kozhevnikov, 1999; Hinton, 1979; Pani, 1997; Pani, Jeffres, Shippey, & Schwartz, 1996). We suggest that when presented with sectional anatomy in the context of knowledge of whole anatomy, participants are able to reason effectively about the mapping between the two spatial domains (consider Gentner & Markman, 1997). Understanding the process of mapping between types of representation and finding ways to better support it is an important area for future research on instruction in advanced disciplines.

The results of this study imply that where there is a significant lapse in time between initial learning of whole anatomy and later instruction of sectional anatomy (e.g., a residency in radiology), students will benefit from a thorough review of whole anatomy in order to optimize transfer. In addition, it is likely that in domains such as histology and microanatomy, in which sections of tissue are viewed through a microscope and good three-dimensional models do not generally exist, a development of three-dimensional models would benefit instruction.

There was no support for the hypothesis that smooth interactive motion through serial anatomical sections facilitates learning of sectional anatomy. A basis for this hypothesis was the common finding that unitary perception of objects follows from continuous sampling of their structure over time. This phenomenon does not appear to occur with serial movement through sectional imagery. In fact, the discrete presentation of sectional anatomy, by focusing attention on single images, led to a small advantage for transfer of knowledge to the interpretation of biomedical images. We suspect, however, that learners would find the continuous navigation more convenient to use. Future work may permit development of a method of instruction that combines the best of both approaches.

Long-term retention of neuroanatomy was high for all groups. This result supports the use of computer-based instruction of anatomy using adaptive exploration. Retention of sectional anatomy was somewhat better for the whole plus sections groups than it was for the sections alone groups. Because the whole plus sections groups required fewer trials with sectional anatomy, this advantage for retention is inconsistent with the standard demonstration of the test effect. In that well-documented procedure, a greater number of tests of knowledge during learning leads to an advantage for long-term retention. However, tests administered during learning and at retention are similar, often identical, to each other (Karpicke & Roediger, 2008). For the present research, such a test effect would show better long-term retention for the sections alone groups than the whole plus sections groups.

On the other hand, the whole plus sections groups did require more total trials to learn neuroanatomy. Thus, the improvement in long-term retention is potentially due to a variation on the test effect. In this case, additional testing of whole anatomy is contributing to long-term retention of sectional anatomy. This is a further instance of transfer of learning and is consistent with the present hypotheses about the benefits of coordinating multiple forms of representation during learning (for related work on transfer and the test effect see Rohrer et al., 2010).

Of course, it may be possible to increase long-term retention in the sections alone groups by increasing the number of learning trials for those groups. These participants would now, however, be required to engage in overlearning beyond an already demanding performance criterion. The overall efficiency of learning would move very much in favor of the whole plus sections groups (for a recent evaluation of overlearning, see Rohrer & Taylor, 2006). This approach also seems to be counterproductive for deployment in actual instruction. In that setting, both students and instructors are inclined to move on from work that has already been presented and tested. If a transfer paradigm can achieve the same retention as overlearning while avoiding that laborious process and teaching two things rather than one, it would seem to be the better method.

Transfer of knowledge to the interpretation of biomedical images was included as a gold-standard test of computer-based learning of neuroanatomy. All groups performed well on those tests, and we consider this to be a strong argument in favor of the present methods of computer-based learning in anatomy. On the other hand, performance in the tests of Uncued Recognition left much room to improve. Changes in instruction that lead to a better understanding of the spatial relations among neural structures and that support the development of spatial reasoning may improve performance in this task. In addition, consider that the interpretation of biomedical images was used in this research only as a test

of transfer of knowledge. Direct instruction using the biomedical images should lead to rapid improvements in recognition performance (see Bransford & Schwartz, 1999).

In summary, the present research was aimed at developing and testing new methods for instruction of neuroanatomy, with benefits for both theory and practice. The project was conceptualized in three areas: psychological theory, instructional strategy, and method of assessment. In the area of theory, neuroanatomy is an intrinsically spatial domain, and this project helped to extend the basic psychology of learning farther into the area of spatial cognition. In particular, it was confirmed that mastery of whole anatomy could serve as the basis for an organization and elaboration of anatomical knowledge that makes learning sectional anatomy more efficient.

The instructional strategy was motivated by the fact that neuroanatomy is a complex spatial domain that is well suited for computer-based instruction. In particular, instruction in this domain depends on high quality graphical presentation and extensive self-study. In an attempt to optimize learning of this type, graphical tools were designed specifically for exploring whole and sectional neuroanatomy and were deployed in the context of spaced practice, repeated testing, and feedback. This combination, which we call adaptive exploration, offers the benefit of unrestricted exploration and discovery in an environment that promotes incremental goal-based learning. The use of realistic materials and procedures was intended to optimize transfer of knowledge and to permit confidence that the methods tested could be deployed in actual instruction.

The method of assessment focused on both short and long term outcomes of the instructional systems that were developed. Thus, the present study included measurement of the efficiency of learning, long-term retention, and transfer of knowledge to interpreting new materials. Overall, this project has provided a prototype for computer-based instruction in neuroanatomy. The outcome of the research encourages the belief that adaptive exploration in computer-based instruction will provide efficient and robust learning in neuroanatomy and similar disciplines.

References

Ackerman, M. J. (1995, October). Accessing the Visible Human Project. *D-Lib Magazine, 1*(4). Retrieved from http://www.dlib.org/dlib/october95/10ackerman.html

Barab, S. (2006). Design-based research: A methodological toolkit for the learning scientist. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 153–170). New York, NY: Cambridge University Press.

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin, 128,* 612–637. doi:10.1037/0033-2909.128.4.612

Bear, M. F., Connors, B. W., & Paradiso, M. A. (2007). *Neuroscience: Exploring the brain* (3rd ed.). New York, NY: Lippincott, Williams, & Wilkins.

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

Bower, G. H., Clark, M. C., Lesgold, A. M., & Winzenz, D. (1969). Hierarchical retrieval schemes in recall of categorized word lists. *Journal of Verbal Learning and Verbal Behavior, 8,* 323–343. doi:10.1016/S0022-5371(69)80124-6

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (2000). *How people learn.* Washington, DC: National Academy Press.

Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. *Review of Research in Education, 24,* 61–100.

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. doi:10.1037/0096-3445.120.3.278

Byrne, R. M. J., & Johnson-Laird, P. N. (1989). Spatial reasoning. *Journal of Memory and Language, 28,* 564–575. doi:10.1016/0749-596X(89)90013-2

Crowley, R. S., Naus, G. J., Stewart, J., & Friedman, C. P. (2003). Development of visual diagnostic expertise in pathology: An information processing study. *Journal of the American Medical Informatics Association, 10,* 39–51. doi:10.1197/jamia.M1123

de Jong, T., Ainsworth, S. E., Dobson, M., van der Hulst, A., Levonen, J., Reimann, P., . . . Swaak, J. (1998). Acquiring knowledge in science and mathematics: The use of multiple representations in technology-based learning environments. In M. W. van Someren, P. Reimann, H. P. A. Boshuizen, & T. de Jong (Eds.), *Learning with multiple representations* (pp. 9–40). Oxford, England: Pergamon.

Duvernoy, H. M. (1999). *The human brain: Surface, blood supply, and three-dimensional anatomy.* New York, NY: Springer.

Garg, A. X., Norman, G. R., Eva, K. W., Spero, L., & Sharan, S. (2002). Is there any real virtue of virtual reality? The minor role of multiple orientations in learning anatomy from computers. *Academic Medicine, 77,* S97–S99. doi:10.1097/00001888-200210001-00030

Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist, 52,* 45–56. doi:10.1037/0003-066X.52.1.45

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology, 12,* 306–355. doi:10.1016/0010-0285(80)90013-4

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15,* 1–38. doi:10.1016/0010-0285(83)90002-6

Graesser, A. C., Jackson, G. T., & McDaniel, B. (2007). AutoTutor holds conversations with learners that are responsive to their cognitive and emotional states. *Educational Technology, 47,* 19–22.

Hegarty, M. (1992). Mental animation: Inferring motion from static displays of mechanical systems. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 1084–1102. doi:10.1037/0278-7393.18.5.1084

Hegarty, M., & Kozhevnikov, M. (1999). Spatial abilities, working memory, and mechanical reasoning. In J. S. Gero & B. Tversky (Eds.), *Visual and spatial reasoning in design.* Retrieved from http://faculty.arch.usyd.edu.au/kcdc/books/VR99/Hegarty.html

Hinton, G. (1979). Some demonstrations of the effects of structural descriptions in mental imagery. *Cognitive Science, 3,* 231–250. doi:10.1207/s15516709cog0303_3

Hohne, K. H., Petersik, A., Pflesser, B., Pommert, A., Priesmeyer, K., Riemer, M., . . . Morris, J. (2003). Voxel-Man 3D navigator: Brain and skull [Computer software]. New York, NY: Springer.

Johnson, C. I., & Mayer, R. E. (2009). A testing effect with multimedia learning. *Journal of Educational Psychology, 101,* 621–629. doi:10.1037/a0015183

Kaiser, M. K., Proffitt, D., Whelan, S., & Hecht, H. (1992). Influence of animation on dynamical judgments. *Journal of Experimental Psychology: Human perception and performance, 18,* 669–689. doi:10.1037/0096-1523.18.3.669

Karpicke, J. D., & Roediger, H. L. (2008, February). The critical importance of retrieval for learning. *Science, 319,* 966–968. doi:10.1126/science.1152408

Keehner, M., Hegarty, M., Cohen, C., Khooshabeh, P., & Montello, D. R. (2008). Spatial reasoning with external visualizations: What matters is what you see, not whether you interact. *Cognitive Science, 32,* 1099–1132. doi:10.1080/03640210801898177

Kikinis, R., Shenton, M. E., Iosifescu, D. V., McCarley, R. W., Saivioronporn, P., Hokama, H. H., . . . Jolesz, F. A. (1996). A digital brain atlas

for surgical planning, model driven segmentation and teaching. *IEEE Transactions on Visualization and Computer Graphics, 2,* 232–241. doi:10.1109/2945.537306

Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction. *Psychological Science, 15,* 661–667. doi:10.1111/j.0956-7976.2004.00737.x

Koedinger, K. R., & Corbett, A. (2006). Cognitive tutors: Technology bringing learning sciences to the classroom. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 61–77). New York, NY: Cambridge University Press.

Kolodner, J. L., & Leake, D. B. (1996). A tutorial introduction to case-based reasoning. In D. B. Leake (Ed.), *Case-based reasoning: Experiences, lessons, and future directions* (pp. 31–65). Cambridge, MA: AAAI Press/MIT Press.

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the "enemy of induction"? *Psychological Science, 19,* 585–592. doi:10.1111/j.1467-9280.2008.02127.x

Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 989–998. doi:10.1037/a0015729

Lesgold, A., Rubinson, H., Feltovich, P., Glaser, R., Klopfer, D., & Wang, Y. (1988). Expertise in a complex skill: Diagnosing x-ray pictures. In R. Glaser, M. T. H., Chi, & M. J. Farr (Eds.), *The nature of expertise* (pp. 311–342). Hillsdale, NJ: Erlbaum.

Levinson, A. J., Weaver, B., Garside, S., McGinn, H., & Norman, G. R. (2007). Virtual reality and brain anatomy: A randomized trial of e-learning instructional designs. *Medical Education, 41,* 495–501. doi:10.1111/j.1365-2929.2006.02694.x

Mai, J. K., Paxinos, G., & Voss, T. (2008). *Atlas of the human brain* (3rd ed.). New York, NY: Elsevier.

Mayer, R. E., Hegarty, M., Mayer, S., & Campbell, J. (2005). When static media promote active learning: Annotated illustrations versus narrated animations in multimedia instruction. *Journal of Experimental Psychology: Applied, 11,* 256–265. doi:10.1037/1076-898X.11.4.256

McNamara, D. S. (2006). Bringing cognitive science into education, and back again: The value of interdisciplinary research. *Cognitive Science, 30,* 605–608. doi:10.1207/s15516709cog0000_77

Melton, A. W. (1970). The situation with respect to the spacing of repetitions and memory. *Journal of Verbal Learning and Verbal Behavior, 9,* 596–606. doi:10.1016/S0022-5371(70)80107-4

Nelson, T. O., & Leonesio, R. J. (1988). Allocation of self-paced study time and the "labor-in-vain effect." *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 676–686. doi:10.1037/0278-7393.14.4.676

Nodine, C. F., & Kundel, H. L. (1987). The cognitive side of visual search in radiology. In J. K. O'Regan & A. Levy-Schoen (Eds.), *Eye movements: From physiology to cognition* (pp. 573–582). Amsterdam, the Netherlands: Elsevier.

Nungester, R. J., & Duchastel, P. C. (1982). Testing versus review: Effects on retention. *Journal of Educational Psychology, 74,* 18–22. doi:10.1037/0022-0663.74.1.18

Paivio, A. (1969). Mental imagery in associative learning and memory. *Psychological Review, 76,* 241–263. doi:10.1037/h0027272

Palmer, S. E. (2002). *Vision science: Photons to phenomenology.* Cambridge, MA: MIT Press.

Pani, J. R. (1997). Descriptions of orientation in physical reasoning. *Current Directions in Psychological Science, 6,* 121–126. doi:10.1111/1467-8721.ep10772884

Pani, J. R., Chariker, J. H., Dawson, T. E., & Johnson, N. (2005). Acquiring new spatial intuitions: Learning to reason about rotations. *Cognitive Psychology, 51,* 285–333. doi:10.1016/j.cogpsych.2005.06.002

Pani, J. R., Chariker, J. H., & Fell, R. D. (2005). Visual cognition in microscopy. *Cogsci 2005: Proceedings of the XXVII Annual Conference of the Cognitive Science Society, 27,* 1702–1707.

Pani, J. R., Jeffres, J. A., Shippey, G., & Schwartz, K. (1996). Imagining projective transformations: Aligned orientations in spatial organization. *Cognitive Psychology, 31,* 125–167. doi:10.1006/cogp.1996.0015

Ratiu, P., Hillen, B., Glaser, J., & Jenkins, D. P. (2003). Visible Human 2.0: The next generation. In J. D. Westwood, H. M. Hoffman, G. T. Mogel, R. Phillips, R. A. Robb, & D. Stredney (Eds.), *Medicine meets virtual reality 11—NextMed: Health horizon* (pp. 275–281). Amsterdam, the Netherlands: IOS Press.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods.* Thousand Oaks, CA: Sage.

Reeves, R. E., Aschenbrenner, J. E., Wordinger, R. J., Roque, R. S., & Sheedlo, H. J. (2004). Improved dissection efficiency in the human gross anatomy laboratory by the integration of computers and modern technology. *Clinical Anatomy, 17,* 337–344. doi:10.1002/ca.10245

Rice, N. (2001). Binomial regression. In A. M. Leyland & H. Goldstein (Eds.), *Multilevel modelling of health statistics* (pp. 27–43). Chichester, England: Wiley.

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning. *Psychological Science, 17,* 249–255. doi:10.1111/j.1467-9280.2006.01693.x

Rohrer, D., & Taylor, K. (2006). The effects of overlearning and distributed practice on the retention of mathematics knowledge. *Applied Cognitive Psychology, 20,* 1209–1224. doi:10.1002/acp.1266

Rohrer, D., & Taylor, K. (2007). The shuffling of mathematics practice problems boosts learning. *Instructional Science, 35,* 481–498. doi:10.1007/s11251-007-9015-8

Rohrer, D., Taylor, K., & Sholar, B. (2010). Tests enhance the transfer of learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 233–239. doi:10.1037/a0017678

Ross, B. H. (1984). Remindings and their effects in learning a cognitive skill. *Cognitive Psychology, 16,* 371–416. doi:10.1016/0010-0285(84)90014-8

Schwartz, D. L., Blair, K. P., Biswas, G., Leelawong, K., & Davis, J. (2007). Animations of thought: Interactivity in the teachable agent paradigm. In R. Lowe & W. Schnotz (Eds.), *Learning with animation: Research and implications for design* (pp. 114–140). Cambridge, England: Cambridge University Press.

Schwartz, D. L., & Heiser, J. (2006). Spatial representations and imagery in learning. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 283–298). New York, NY: Cambridge University Press.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence.* New York, NY: Oxford University Press.

Tulving, E. (1962). Subjective organization in free recall of "unrelated" words. *Psychological Review, 69,* 344–354. doi:10.1037/h0043150

Tversky, B., Morrison, J. B., & Betrancourt, M. (2002). Animation: Can it facilitate? *International Journal of Human-Computer Studies, 57,* 247–262. doi:10.1006/ijhc.2002.1017

van der Meij, J., & de Jong, T. (2004). *Learning with multiple representations: Supporting students' translation between representations in a simulation-based learning environment.* Paper presented at the American Educational Research Association, San Diego, CA.

van Someren, M. W., Reimann, P., Boshuizen, H. P. A., & de Jong, T. (1998). *Learning with multiple representations.* Oxford, England: Elsevier Science.

Zimmerman, J. (1975). Free recall after self-paced study: A test of the attention explanation of the spacing effect. *American Journal of Psychology, 88,* 277–291. doi:10.2307/1421597