Exploring a Cognitive Basis for Learning Spatial Relationships with Augmented Reality

BRETT E. SHELTON[1]* AND NICHOLAS R. HEDLEY[2]

[1] *Department of Instructional Technology, Utah State University, Logan, UT, USA*
[2] *Department of Geography, University of Washington, Seattle, WA, USA*

Augmented reality (AR) is an emergent class of interface that presents compelling possibilities for advancing spatial visualization. We offer a brief overview of AR technology and current research with in the educational realm. AR interfaces appear to provide a unique combination of visual display properties, modes of user manipulation, and interaction with spatial information. Drawing upon aspects of proprioception and sensorimotor function, we discuss how AR may have a unique and powerful link to spatial knowledge acquisition through visuo-motor involvement in the processing of information. We identify key properties of AR interfaces and how they differ from conventional visualization interfaces, followed by a discussion of theoretical perspectives that make a case for learning spatial relationships using 1st person manipulative AR. Recent research provides evidence that this form of AR holds cognitive advantages for learning when compared with traditional desktop 2D interfaces. We review the visual-physical connections to learning using 1st person manipulative AR within educational contexts. We then provide some suggestions for building future research in this area and explore its significance in the realm of spatial knowledge acquisition.

Keywords: Augmented reality, spatial visualization, animate vision, learning theory.

1. INTRODUCTION

This paper is the result of an ongoing collaboration between the authors that resulted in the consideration of cognitive constructs based on knowledge acquisition using advanced spatial visualization tools – specifically

---
Corresponding Author: e-mail: brett.shelton@usu.edu; nix@u.washington.edu

FIGURE 1
A view of a student interacting with real objects (foamcore card, table, wall) and artificial
objects (Sun, Earths, annotations) through the Augmented Reality interface. This view is as
would be seen if wearing an HMD.

augmented reality (AR) interfaces. AR interfaces mix real views with virtu-
al objects, allowing users to view 3D visualizations in familiar everyday set-
tings, without the disorientation and expense of traditional immersed virtu-
al environments (VE). This is achieved by wearing a liquid crystal display
that provides views of the real world enhanced with virtual content (see
Figure 1). The unique characteristics of 1st person manipulative AR appear
to embody significant potential for the cognition of visualizations of spatial
information.

AR can display representations of spatial phenomena in a way that uti-
lizes 3D visualization technology while avoiding the drawbacks of com-
pletely immersed systems that occlude the real world. From a spatial knowl-
edge acquisition and cognitive perspective, AR may utilize unique cognitive
mechanisms for spatial knowledge acquisition. This has significant implica-
tions for spatial knowledge acquisition and interface design principles.

First we provide an introduction to augmented reality, including technol-
ogy, implementation and use. We discuss the practical significance of AR in
the context of spatial visualization, and the implications for spatial cogni-
tion. It is followed by an overview of AR in education, and an introduction
to theory for understanding how people learn during interactions with AR
interfaces. We broadly consider a cognitive basis for AR use emphasizing
visuo-motor involvement in the processing of information. First-person
AR's unique properties suggest that revisiting and integrating theories orig-
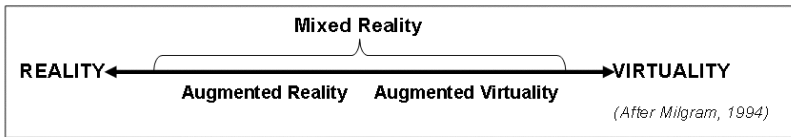inating from different disciplines might be useful. We draw upon aspects of

FIGURE 2
Augmented Reality and Mixed Reality on the Reality-Virtuality spectrum proposed by Milgram (1994).

proprioception and sensorimotor function from research in artificial intelligence in discussing visual knowledge acquisition, and ways that the AR interface may exploit certain advantages given its unique properties. We discuss recent evidence from studies using spatial referents as a rubric for its effectiveness. We present a case where 1st person manipulative AR was used to teach undergraduate students Earth-Sun relationships, paying special attention to interface and visuomotor learning activity. Finally, we discuss areas of future research and its significance for spatial visualizations.

2. AUGMENTED REALITY (AR)

AR interfaces enhance reality by mixing real views with virtual objects (Azuma, 1997). Milgram (1994) described "mixed reality" as occupying the middle ground between reality and virtuality (see Figure 2). There are several permutations of AR, including "annotated vision" in which real views are enhanced through real-time annotation. Other AR systems include novel applications where users "fly into" immersive virtual worlds and participants can leave their real surroundings behind, and join others to collaborate in shared virtual spaces (Billinghurst, Kato, & Poupyrev, 2001).

So how does 1st person manipulative AR work? The interface software is run from a single-workstation computer and is used by a wearing lightweight liquid crystal head-mounted display (HMD). The user views handheld cards (platform), to which virtual objects are rendered and oriented in 3D space by the software. As the user moves the card, the virtual object stays anchored to the card and moves as if attached to it. The resulting effect is simple, yet dramatic. Previously mouse or button-actuated, and metaphor-mediated activities—such as zoom, pan and rotate—are now achieved by moving the card in one's hands. This is a familiar, intuitive activity that does not require any specialized understanding.

FIGURE 3
Using 3D molecular models with AR for learning biology and chemistry (HITLab, 2002).

   AR can be traced back to Sutherland's work using half-silvered optics in
the 1960s. It has only been recently that AR has really matured and become
robust enough to begin using it in applied contexts. The past couple of years
in particular have seen the popularity of AR spill over into the popular media
and public interest (Ditlea, 2003; Feiner, 2002). Before then, most of AR
research effort has gone into establishing more and more stable hardware, in
addition to improving computer vision and registration algorithms (Azuma,
1997; Azuma et al., 2001). Far less work has been done to investigate and
develop the interface and its applications beyond laboratory settings. User
studies have concentrated on low-level perception (such as depth perception
and latency effects), rather than higher-level knowledge acquisition studies
(Drascic & Milgram, 1996; Neumann & Majoros, 1998; Rolland & Fuchs,
2000). This recently achieved robustness presents an exciting challenge.

3. USING AR FOR EDUCATION

   There is a vast body of research that deals with fully immersive virtual
reality (VR) and education beginning in the early 1990s (e.g., Dede, 1995;
Osberg, 1993; Winn, 1995) continuing through recent research efforts (e.g.,
Barab, Hay, Barnett, & Keating, 2000; Winn, 2002). AR research has lagged
somewhat behind VR because of its delay in technological developments
and creating practical applications for its use. The most recent AR research
within the educational realm is looking at topics that naturally lend them-
selves to 3D space, such as the 3D structure of molecules (HITLab, 2002).
The HITLab, in conjunction the Scripps Research Institute and the

FIGURE 4
Using drag-and-drop to create new combinations of molecules (Fjeld, Schar, Signorello, & Krueger, 2002).

University of Utah, has initiated a research project that teaches molecular biology concepts to high school students. Teachers and students experiment with different kinds of 3D molecular models and discover new ways of interacting with them. Instructors at a Seattle high school are working with the research team to develop lessons that may be taught using AR technology. So far, the response from all parties has been positive. The high school will use AR for teaching biology and chemistry this winter, with plans to expand the program to include more complex concepts and techniques in the near future (see Figure 3).

Developers and researchers in Switzerland have created a kind of AR virtual chemistry laboratory (see Figure 4). Students can view and acquire simple atoms through a virtual drag-and-drop technique. Atoms get combined by matching the spinning outermost electrons of a particular atom to ones that fill its required shell. Once combined, a new structure is seen and additional atoms can be added using the same method as before. Labels that give the name of the structure appear when "completed" molecules are formed. This way students can construct their own complex molecules while being bound by the subatomic rules that govern molecular interactions. This feature offers a clear advantage over traditional methods of building models using styrofoam and straws (Fjeld, Schar, Signorello, & Krueger, 2002).

There are potential educational applications besides those in science and engineering. Researchers at the University of Singapore have developed a system that uses motion capturing, that when applied to computer models, mimics the actions of dancers as they perform (Cheok et al., 2002). When viewed through an AR system, students and instructors can experience the performance from any angle as many times as they want in order to make modifications to a scene, critique actions, or simply enjoy it. Seen here, vir-

FIGURE 5
Virtual dancers performing life-size on a rooftop (Cheok et al., 2002).

tual dancers perform life-sized on a rooftop (see Figure 5).

Therefore, as the use of different forms of AR become more widespread, making efforts to understand the underlying cognitive processes become increasingly important. The 3D molecule project uses a 1st person perspective with tracking built-in to physical 3D models. The AR chemistry lab uses 3rd person perspectives, viewing action on a mirrored display. Viewing the dancers from Singapore relies on 1st person perspectives without any opportunities to physically change perspectives with the dancers themselves through handling of the visual marker. But which affordances embedded within the different forms of AR are most advantageous for learning? How do people acquire knowledge about spatial phenomena based on interactions with these interfaces? In the following sections, we revisit theory that may be integrated to inform AR interface design and use. These are: spatial cognition theory; animate vision; and vision theory in education.

4. THE CASE FOR 1ST PERSON AR WITH ALLOWANCES FOR PHYSICAL INTERACTION

Spatial cognition theory

From a spatial cognition perspective, the AR interface raises some interesting questions. When using AR, one is no longer detached from the 3D

FIGURE 6
Collaborative use of 3D landscape visualization using AR (Hedley & Shelton, 2002).

content through desktop metaphors and the inherent ambiguity between mouse clicks and buttons on a screen. People acquire knowledge about spatial phenomena, by viewing 3D objects such as landscapes held in their hands (see Figure 6). Visual and spatial queries take the form of everyday object manipulation. In other words, for you to see the other side of an object, you rotate it in your hands. To see more detail, you bring it closer to your eyes.

What kind of spatial knowledge does an AR interface provide? In the spatial cognition literature, three *types* of spatial knowledge are often mentioned: (1) *procedural knowledge* - that which allows us to get around in geographic space and the information that forms the basis for navigation and wayfinding; (2) *declarative knowledge* - simple facts about geographic space and the entities within it; (3) *configurational knowledge* - knowledge of geographical space that is essentially map-like though it contains information about relative positions, orientations, distances, and relationships between spatial entities (Golledge, 1991; Mark, 1992; Mark & Freundschuh, 1995).

AR interfaces are likely to constitute some combination of procedural or configurational knowledge. It may be procedural due to the fact the some AR interfaces allow the user to "fly into" the 3D display, and experience it as if standing in or moving around inside a virtual world such as the Magic Book (Billinghurst et al., 2001). It may also be configurational due to interaction modalities where a user holds a 3D landscape in her hands like a map,

and views the entire geographical space in one view.

Perhaps AR users are getting a better sense of 3D content because of the cognitive pathways through which spatial knowledge is perceived, verified, triangulated and internalized. Distinctions between types of spatial knowledge by *source* of have been proposed. These types of spaces include: *(1) haptic* - spatial knowledge based on touch or body movement; *(2) pictorial* - spatial knowledge based on information in visual form; *(3) transperceptual* - spatial knowledge based on a combination of multiple information sources and/or experiences synthesized over a period of time (see Mark, 1992; 1993). From the perspective of this typology, AR provides haptic and pictorial spatial knowledge. The haptic spatial knowledge is gained through physical action. This physical action is not locomotion-based, which would be more closely linked to procedural knowledge, rather it is derived from in situ physical manipulation.

A compelling aspect of AR interfaces is that the combination of strong pictorial and haptic spatial knowledge acquired from interaction and manipulation may result in more rapid and more accurate perception. Essentially, multiple reference frames combine to enhance the cognitive experience and transfer of spatial information. This pictorial and haptical spatial knowledge can then be integrated into the individual's working cognitive model of the spatial phenomenon in question.

How might this integration occur? A highly appropriate suggestion comes from Portugali who suggests that there is an interaction between internally stored representations derived from previous environments and the perception of external patterns in the new environment (1996). Portugali's work extends Neisser's (1976) transactional model of cognition, integrating the external world (to maximize ecological validity of methods and the importance of studying memory in the world rather than the laboratory). In Portugali's (1996) framework, the comparison of these internal and external representations, or Inter-Representation Networks (IRN), result in a dynamically stable internal representation which forms the basis for interactions with new spatial information sources. These sources might be experiential, map-like or could include verbal information. Together they once again constitute transperceptual spatial knowledge. Integrating varied forms of spatial knowledge with existing knowledge from previous experiences results in what Haken called a synergetic system (1991). A synergetic system is a dynamically stable internal representation–essentially a working understanding–that continuously updates as information and experiences are added to existing ones.

Spatial cognition theory can provide a number of individual and integrated theoretical devices for understanding the multi-modal nature of the AR interface. This in turn provides a mechanism to accommodate the multiple cognitive pathways through which spatial knowledge appears to be acquired in this setting.

AR interfaces hold opportunities for knowledge acquisition, through visual, spatial and sensorimotor feedback. These characteristics embody different mechanisms by which users acquire spatial knowledge. We propose that AR works because of visual and spatial cues set in the context of everyday user surroundings. In addition, we propose that AR is a particularly powerful tool for spatial visualization due to the sensorimotor feedback users receive in response to manipulation inputs combined with visual and spatial cues. In the next section we consider how sensorimotor exploration and vision are tightly linked to the concept contained in animate vision theory.

Animate vision theory

Animate vision is a theory that links visual concept acquisition to acting and moving in the physical world, often not distinguished between "active vision" theory (Aloimonos, 1993; Ballard, 1991). Clark (1997) describes it as a visuo-motor theory – that humans sample a scene from the world in ways suited to their immediate needs. Vision is not the transformation of light signals into a representation of an enveloping 3D world, but instead a series of fast adaptive responses that cycle into routines of acting and moving within an environment. The crux of animate vision theory and others related to it (such as inattentional blindness) may be that vision is a tool used for sensory exploration of the environment, using an action-involving cycle of fragmentary perception, similar to the sense of touch outlined by Mackay (1967) and O'Regan (1992). Other experiments that involve altering visual displays during saccades offer support for considering vision as a series of interconnected partial representations of the physical world (Rayner, Well, & Pollarsek, 1980). Visual scenes may be nothing more than a kind of "subjective illusion" caused by the continuous scanning of small areas using short attention periods (Clark, 1997).

Studies of visual perception have pointed to an inextricable link between motor movement and the visual system that implies an implicit knowledge representation. Dienes and Perner (1999) contend the visuo-motor system monitors neither time nor facts necessary for explicit knowledge. Understanding that the visual and motor systems are linked with implicit

knowledge is important because it is unconscious information impervious to illusion and cognitive misconception. Correctly deriving information about one's environment, and the physical presence and actions within it, may therefore be paramount in making conscious cognitive assertions of factuality, eventually leading to the encoding of information and decision making.

The initial stages of visual stimuli are mediated by perceptional information that facilitate the identification and discrimination of visual concepts, prior to any cognitive interventions formed by prior exposure to visual cues for discrimination or categorization. This is not to assert that perception cannot occur as a global level. Rather, the processes that govern how visual stimuli are selected and applied are interactive in the way we perceive our environment. Previous research has acknowledged that the size and discriminability of the parts and the density of details in a visual image may be factors in how an object is ultimately processed (Winn, 1994). Here, we suggest that the nature of the visual image cannot be separated from the action of the individual who perceives the image; that iterative processing is governed by visual and motor processes alike.

If this emerging theory of human vision that involves active physical behaviors combined with visually-related behavior is accurate, it is then important to consider how an interface that reflects the visual contribution to learning can be built, applied, and researched for its usefulness.

Vision, research and theory in the context of learning

Perhaps learning with visualizations should take advantage of interfaces that combine affordances of visual stimuli and motor responses. The augmented reality interface lends itself well to task-related learning because of the exclusive connectivity between short cycles of visual perceptual activity and physical movements. This provides the user with advantages for action in the world and physical processes that involve action. Interfaces can help explain how people learn due to their dual visual and physical interactive nature.

Further technology research can use schema theory combined with active vision theory in the use of visualizations for education. First of all, the theories assume that schema can be effectively built and activated through information presentation closely resembling the structure of a particular schema. Two methods provide explanation on how this can be accomplished: (1) direct image encoding as mental image and (2) propositions based on pictorial information that can be reconstructed as a mental image. Other research looks at how this information might be dually coded in both

manners. The advantage of multiple codings is redundancy and uniqueness, being able to recall information in a variety of forms for a particular purpose. Moving elements of the environment may also affect the way a person interprets the intentions of the objects within that environment (Shelton, Humble, & Matson, 1996). In active vision theory, the nature of the visual image cannot be separated from the action of the individual who perceives the image. Ultimately, iterative processing is governed by visual and motor perceptual processes alike. It is important to concentrate on how visualizations are used in the process of learning and how different visual representations are utilized for students of varying levels of prior knowledge and possessing different learning attributes.

Visualization research may also focus on how students impose their own structure on incoming information for more effective learning. Winn and Snyder (1996) refer to this process as "information mapping." Students, by organizing content themselves, may have significant improvement due to spatial presentation and layout with information recall. Results from VR research have provided a basis for recommending layout for graphical and pictorial representations in instructional materials (Winn & Windschitl, 2002).

Sensory exploration and learning

Neisser describes schema as both an information accepting system and as a planning system for finding out about objects and events. Accepted information in one cycle might then be used as part of the system for accepting new information. "By constructing an anticipatory schema, the perceiver engages in an act that involves information from the environment as well as his own cognitive mechanisms" (Neisser, 1976, p. 57). The new information then changes the perceptual schema and might affect the next act of the perceiver.

Cognitive maps are active, information seeking structures that contain spatial imagery, accept information, and direct action. Schemata are embedded in the cyclical system of cognitive maps, specifically attuned to the environment and perceptual exploration. The cognitive map of the world directs locomotion and action; action samples properties of the actual world to acquire potentially available information, and the actual world modifies the cognitive map of the world.

While multiple sensory exploration of an object is valuable for triangulating perceived information, visual perception is most effective during motion. Motion produced information is critical for effective vision, as it

provides valuable information about objects in relation to one another in the
environment and the movement of the perceiver in relation to those objects.
Schema incorporate potential locations for things that have not yet been
seen. Information acquired through motion is linked to existing schema and
to the cognitive map of the environment in particular.

   Information about oneself, like all other information, can only be picked
up by an appropriately tuned schema. Conversely, all information that is
picked up, including proprioceptive information, modifies a schema. In the
case of movement through the environment, this is an orienting schema or
cognitive map. This means that the cognitive map always includes the per-
ceiver as well as the environment. Ego and world are perceptually insepara-
ble (Neisser, 1976).

Cognitive organizations

   A general feature of cognitive organization is that units at lower levels of
abstraction feed information to other higher levels. They are not related
sequentially, but embedded, each engaging in its own cyclical system with
the environment. The mechanisms for knowledge representation exist inside
behavioral processes. "A behavior is a sequence of cognitive events and
actions, a set of visual, planning, memory, and reasoning processes working
in a cooperative manner and 'acting' on the system itself or its environment"
(Aloimonos, 1993, p. 8). Vision's purpose is action. An adaptable and prac-
tical visual system is meaningless without action. Learning is more success-
ful in the active vision model because of its inclination toward well-defined
behaviors instead of general purpose representations set only in theoretical
conjectures.

   Ballard (1991) suggests smaller objects are linked to larger objects in
mental representations. Since the Marr (1982) paradigm was introduced,
research for object identification in context has languished. Working on gen-
eral vision has discouraged the integration of learning and visual processes
(Aloimonos, 1993). By focusing not on the generality of vision, active
vision advocates making the utmost use of all different kinds of constraints
placed on objects in context. An example of environmental constraints on
objects is the vestibular system in humans that measures linear and angular
accelerations. This system provides a brief history of movements in the envi-
ronment as well as a measure of gravitational force. An additional source of
information is the local context of objects themselves, such as the design for
affordances of objects in the way they interact with supporting surfaces. An
example is the way a coffee mug is designed, so that visual recognition tasks

are simplified.

AR appears to be a compelling environment in which to engage spatial phenomena. Other design implications exist that support experience in AR. We can identify two factors, supported by animate vision and visual concept acquisition theory, that provide an integrated understanding of AR interface use.

First, the retention of proprioception. In AR, the participant retains the proprioception of self within the environment. That is, the unconscious awareness of one's own physical presence in space remains intact. Often virtual environments neglect the idea of representing the participant's physical space in the environment, instead relying on a smaller representation as an avatar or glove that "floats" in space without parallel representation of the body of the participant.

Second, retention or sensorimotor function. With AR, the action within the environment is created by physical movements initiated by the participant. The exploratory senses involved in saccadic movement do not conflict with the physical location of the physical 3D space. In addition, other sensorimotor processes of temperature and texture, audio and olfactory, all remain true to the encoding of implicit knowledge. Artificial sensory feedback of the environment such as force-feedback mechanisms in peripheral devices are no longer necessary.

Consideration of these theoretical approaches leads to a multi-perspective theoretical foundation for understanding how AR may operate as an interface at the cognitive level. At the same time, AR has properties that can be understood by integrating previously separate theoretical frameworks. In the following section, we synthesize aspects of the AR interface and theory that we have discussed up to this point.

Synthesis of theory and interface

We have presented an overview of the augmented reality interface and its characteristics and have proposed several theoretical approaches to understanding its use in learning. This theoretical framework draws upon spatial cognition, animate vision and vision theory in the context of learning. In this section we highlight compelling aspects of the AR interface, linking them to theory we have presented above. Following this we provide an integrated view of these perspectives and the relationship to AR use in educational settings.

A key advantage of AR is that participants retain their view of the world while interacting with virtual objects. From a research perspective, using
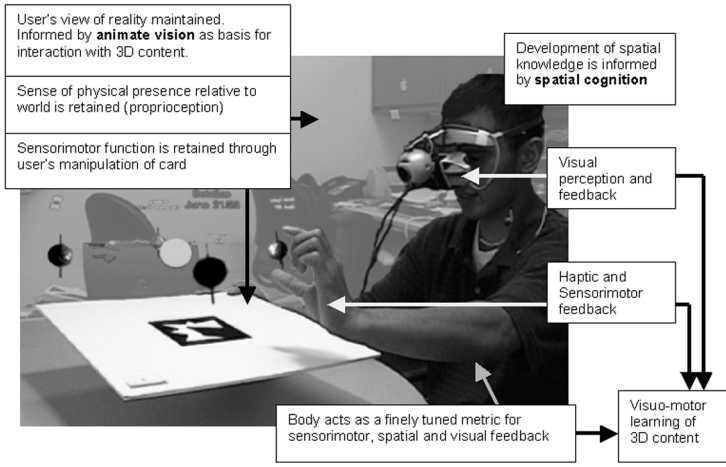
FIGURE 7
Proposed theoretical mechanisms of the AR interface (Shelton & Hedley, 2001).

virtual content allows total control over content as stimuli in experimental design. Unlike immersive virtual environments, AR interfaces present animated virtual content at the same time as maintaining real-world surroundings. AR provides the benefits of controlled content for research while placing it in a familiar ecological context.

One of the disadvantages of a completely immersive virtual environment is the loss of "sense of self" in space. That is, in immersive virtual reality one's self is composed of small components of the physical self through the use of avatars or portions of one's body, acting in a space consisting of virtual objects. This has often resulted in feelings of disorientation and difficulty in movement and intended behaviors in the virtual environment. Animate vision theory describes the importance of maintaining a sense of physical presence in order to make behaviors that are based upon information obtained moment-to-moment. In augmented reality, the person is able to combine the 3D object into the normal viewing perspective without losing any of the advantages of object movement and individual movement that creates the behaviors that help us perform activity (gain sensorial-based knowledge) in real-world environments.

Why can't this be done through conventional manipulation in computer terminal virtual object manipulation? In virtual environments experienced through a computer desktop-based application, manipulation of an object is
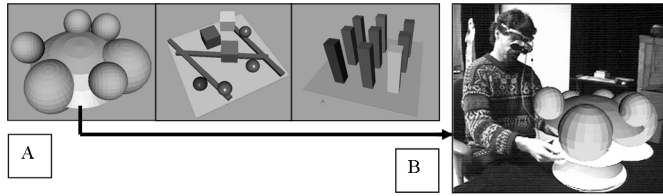
FIGURE 8
Examples of content and interface use in intial emperical cognitive tests of AR use. A: 3D models. B: participant using models in experimental setting (Hedley 2001).

performed through operations that "filter" the manipulation of an object. In other words, in order to rotate or move a virtual object, the person must cognitively "transform" these operations into 1) move mouse cursor over appropriate button, 2) click button, 3) see object orientation change, and 4) process the result in order to create additional mouse clicks. This brief list is greatly simplified to explain the complicated cognitive and motor processes needed in order to make a desktop virtual object change its orientation.

The point is that such processes may inhibit the acquisition of visual information. Animate vision theory advocates the direct physical manipulation of an object for the effective computation of object recognition—and eventual understandings in accordance with this recognition. This "filtering" effect of desktop interfaces may inhibit the effective cognitive processes involved in assimilating and accommodating information (Hedley, 2001).

Having laid out a number of theoretical propositions in preceding sections, and having highlighted important features above, we can revisit the first view we provided of AR being used in this paper (Figure 1), and indicate where the components of our cognitive theory fit into real-world AR interface use. These associations are presented in Figure 7.

5. EMPIRICAL EVIDENCE OF 1ST PERSON PERSPECTIVE MANIPULATIVE AR

AR and spatial referents

Researchers at the University of Washington Human Interface Technology Laboratory have conducted a number of experiments to gather empirical evidence as a basis for theoretical propositions and validation. Many informal assessments (at least 50 individuals) have resulted in a strong sense that AR is a powerful and engaging visual and cognitive experience
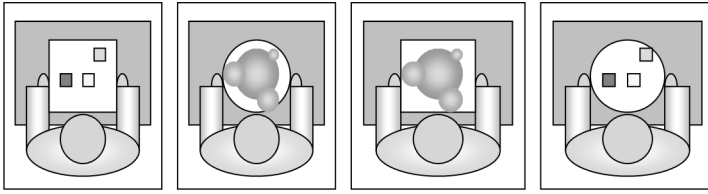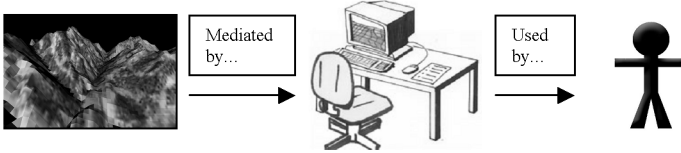
FIGURE 9
Content-platform combinations used in early experiments. Combinations of square and circular cards with orthogonal or irregular content were used to identify the role of visual and spatial cues gained from the card itself versus the 3D model content (Hedley, 2001; 2003a).

FIGURE 10
Identical experimental activities, varying only desktop (DT) and augmented reality (AR) interfaces between treatments (Hedley, 2003a; 2003b).

for users. In initial experiments, 20 participants performed a range of tasks involving basic 3D visualizations. These tasks were aimed at understanding the role of visual and spatial cues embedded in: the visualization (Figure 8A); the physical interface (card) (Figure 8B); and the user setting (Figure 8B and Figure 9) of AR interfaces (Hedley, 2001).

Experiments included the manipulation of the 3D content, such as linear and planar objects versus curved and spheroid objects (Figure 8A). Tasks required participants to judge relative distances between objects in the visualization and rotate the 3D models to specified angles relative to the staring orientation. In some treatments, subjects were not allowed to touch or move cards while in others different combinations of regular versus irregular content were matched with square or circular cards (Figure 9). This latter treatment yielded insights into users' use of visual and spatial cues from the cards and or 3D models in order to complete tasks.

Some of the interesting learning activities that emerged were the strategies used by participants to perform tasks. There were distinct combinations of visual and spatial cues used by people with different training. Individuals demonstrated different prototyping strategies to complete tasks. During rotation tasks, there were distinct decreases in performance of participants when primary spatial referents were removed or reduced. Subjects were more accurate in determining distances between objects in 3D models when corners, linear features and orthogonal structures were present in either the interface (table or card) or in the visualization. In the absence of these visual and spatial referents, some individuals used their hands as referents and metrics. This strategy was also seen in distance estimation tasks between objects in the 3D models. Participants often looked for unique landmark objects to use as spatial referents. In the absence of unique features, many resorted to a "body as referent" tactic once more. The results support the notion that people learn relative spatial relationships by using perceived referents during physical manipulation of virtual objects.

AR, perception, manipulation, and performance

Subsequent to this early experiment, Hedley (2003a; 2003b) designed and undertook a 100-person designed an empirical experiment to study between-groups differences in performance, behavior and cognitive maps due to the mediating effects of desktop versus augmented reality interfaces. This was supplemented by a within-treatments analysis of the influence of visualization content and user characteristics on cognitive representations.

Participants were required to engage in perceptual tasks, solve spatial decision-making tasks, and memorize visualizations. Two groups of 50 participants (100 total) engaged in identical experimental activities. The only difference between treatments was that one group used a desktop interface to interact with 3D visualization content, while the other group used an AR interface to interact with the content. The behavior, performance and task responses of individuals using an AR geographic visualization interface were compared to identical activities undertaken by users of a desktop 3D geographic visualization interface (see Figure 10). An array of quantitative and qualitative data were gathered from 101 participants during 250 hours of observation.

Visualization content and interface were predefined, controlled and manipulated during these treatments. User training and spatial ability were measured. Participants' perception, judgment and internalization responses were measured during experimental activities, providing data about com-

pleteness and detail of users' internal representations, speed and accuracy of timed and untimed spatial perception and problem-solving tasks.

Quantitative analysis focused on relationships identified between interface characteristics, visualization content, user characteristics, and the resulting cognitive representations, judgment and performance observed and measured during user activities. Cognitive representations were measured and evaluated in terms of completeness and level of detail. Users' performance and judgment in spatial perception and problem-solving tasks were also evaluated. In all cases, the data were gathered during or in response to a set of five visualization stimuli, and also during timed problem-solving tasks immediately preceding and following the sequence of 3D model interactions. Preceding analysis, tests established that there were no confounding differences in user experience or training between interface treatment groups that may have had an influence on other tests in the analysis.

AR interface use resulted in significant beneficial influence on perception, performance and inferred cognitive representation of 3D geographic visualizations. Significant differences in the distributions of users' representation scores showed that AR interface users' minimum scores in spatial visualization tasks were higher than the equivalent activities performed using desktop interfaces. Significant results were also found in investigating the role of visualization content and physical manipulation of the AR interface, over the use of desktop interface.

AR users consistently produced greater levels of completeness in representations. This displacement was clearly visible in the analysis of cumulative frequency distributions. In addition to this observation, the minimum number of features by AR users was represented consistently higher than the minimum performance exhibited by desktop interface users. Lorenz curve analysis provided a visual means to study the quantitative difference in equality of distributions, and showed the relative inequality of feature score distributions and where the majority of feature score performance existed for each interface type, within the respective distribution. AR interface users were seen to have less balanced performance across the population, with performance being weighted towards higher performance than desktop users. Finally, regression analysis found AR interface use significantly predicted higher feature score performance. The advantages found by AR users over desktop users were attributed to the multisensory interactions AR interfaces provide. Direct manipulation of cards augmented with virtual content provides a more transparent interface (one with few layers of metaphor, etc.). At the same time, the coupling of visuo-motor feedback and proprio-

ception provided a powerful sense-making experience, grounded with a stable frame of reference. AR use resulted in higher levels of detail in representations than desktop interface use. More AR users produced higher level of detail representations than desktop users, and differences were found to be significant for two of the stimulus cases. A significant positive correlation was found between AR use and higher levels of detail in representations for three of the stimulus cases. AR use was found to significantly predict higher levels of detail sixty per cent of stimulus cases.

In a standardized spatial problem-solving activity repeated at the start and end of the experiment, AR users were seen to accurately complete spatial problem-solving tasks 1.57 seconds (22%) faster than desktop users at the start of the experiment, while this margin reduced to 0.3 seconds (5%) percent faster than desktop users when the activity was repeated at the end of the experiment. While this was not a familiar spatial problem-solving task, the idea was to use 3D model content of sufficient abstractness for it to be unfamiliar to all users. By minimizing the potential confounding effects (such as specialized training) on a measure of cumulative exposure to the interface, it allows more certainty concerning the results of the primary visualization.

The difference in speed of accurate response suggests that, all other things being equal, AR interfaces have less cognitive inertia than desktop interfaces. That is, in an unprimed setting, user X will be able to understand and interact with content via the interface faster than with a desktop interface. If this is the case, it could be important for situations where user expertise cannot be assumed (such as museums and educational settings), and in other situations where maximum speed of content internalization and task performance is critical (such as air traffic control and strategic decision making).

Cognitive load suggests that learning happens best under conditions that are aligned with human cognitive architecture, and aims to achieve this through the evaluation and design of learning practice and technologies, among other things. In this instance, performance for less adept participants may be due to the multisensory nature of the AR interface. That is – the cognitive load is spread across multiple sensory pathways. This does not guarantee better performance, but may maximize the potential of different users' cognitive architecture. The second finding noted above identifies that performance increased longitudinally through the experiment, with time. This training effect is to be expected, as one might expect that with repeated use, user familiarity and skill with interface, content and protocols might

increase. However, it appears that the effect was not significantly different between interface groups. This reinforces the previous finding about inertia.

In an assessment of interface manipulation on accuracy of spatial judgments, AR users achieved higher levels of accuracy in static and dynamic spatial tasks than desktop users. However, when the differences between AR and desktop performances were considered in each of static and dynamic treatment, desktop user performance converged with AR user performance. The relationships between 3D model content and static or dynamic manipulation treatment showed AR performance to be higher in both. While using the same 3D model stimulus, the combination of desktop interface and manipulation a stronger effect than for AR and manipulation. The desktop score converged with the AR score in the dynamic treatment. The fact that the desktop score converged with the AR score in the manipulation experiment suggests that manipulation is a more important variable than simply changing from desktop to AR. This finding supports previous findings in studies of virtual environments, where interaction was found to be a more important variable than immersion (Byrne, 1996).

In addition to interface convergence, manipulation produced a larger increase in accuracy when the visualization content was mismatched with structural content, versus when it was matched. This suggests that manipulation activates 3D visualizations used in desktop interface settings. Movement of the 3D model on the display will help users make sense of the 2D image, which provides the illusion of displaying a 3D model. The influence of physical manipulation was seen to be a significant factor in determining completeness of representations. The evidence suggests that the activation of visualization-based performance by manipulation is due to the operation of emergent symbol systems embedded in the human-visualization relationship.

Manipulated mismatches between thematic and structural information influenced the level of detail and features scores of subjects. It appeared that the 'cognitive signal' in the mismatched thematic model was being interfered with – either through misinformation or confusion or uncertainty resulting from conflicting evidence in participants' perceptions of the 3D model.

Other user factors observed during the experiments include unique experience-based effects. For example, during the use of a digital elevation model (DEM) of Honolulu, one subject provided a highly detailed description of the coastline, including numbers of inlets, detailed descriptions of the shape of inlets. His description of the areas away from the shore were unre-

markable, if not poor compared with other participants. The post-test interview provided an opportunity to try and unpack not just an emphasis on the coastline, but a sophisticated approach to detailed representation. As it turned out, the subject was an amateur sailor. This reveals the possibility for highly-tuned active search and perception of meaningful features of interest or for action. This observation is informed by the work of Lowe (1993), who compared the ways in which meteorologists construct mental representations from weather maps. Distinct differences were found between the performance of professional meteorologists and non-meteorologists. Non-meteorologists were found to focus on superficial, domain-general, visuo-spatial features, and could recognize spatial patterns in the diagram but were not adept at translating this spatial knowledge into weather knowledge. Professional meteorologists were more skillful at selecting those visual features that are essential for developing an understanding the state of the weather system being depicted, and were better able to decode the semantic analogies — between the visuo-spatial characteristics of the diagram and the physical characteristics of the weather system encoded into the maps.

Spatial ability (measured by Vandenberg MRT) and AR interface use in the experimental activity were the most important predictors of more complete, detailed and accurate responses. Spatial ability and AR interface use were the most significant predictors of higher levels of representations of detail and task performance. When the interface was not carrying the cognitive load, the user's spatial ability compensated, and when spatial ability was not useful, the interface did the work. Visualization experience and spatial training were next in importance. This suggests that the interface and spatial ability may be more influential on spatial knowledge acquisition from geographic visualization interfaces than spatial training or visualization experience. There may be a symbiotic relationship here – spatial ability in users is activated by the inherently spatial AR interface, whereas the desktop interface operates through understanding a set of controls, metaphors and symbols which result in rotation and manipulation of 3D visualizations.

AR and learning Earth-Sun relationships

Based on the research just described that linked the benefits of learning through 1st person manipulative AR, we looked to test its effectiveness in a school context. We built a learning activity using AR to teach Earth-Sun relationships to undergraduate geography students. The results indicated trends useful for application toward spatial learning (Shelton & Hedley,

2002). We found that visual and physical activity positively impacted the learning of dynamic spatial relationships.

The students who were physically active, changing the 3D orientation and position of the instructional content, tended to display a high level of understanding when questioned about their knowledge of Earth-Sun relationships. Conversely, students who viewed the 3D content from fewer perspectives displayed less complete understandings. The more the student interacted with the virtual objects, the more they seemed to learn. This interaction was characterized by rapid and continuous movement of the card as a result of student and instructor-driven tasks for exploring the objects—and meaning behind—what they were seeing.

The following sections support the claim that visuo-motor activity is linked to the students' learning of dynamic spatial relationships through Shelton's analysis (2003) of students' videotaped learning activities. A total of 43 undergraduate students took part in the learning exercise that followed a form of pre-test, AR activity, post-test. The students were tested on their knowledge about Earth-Sun relationships in the form of rotation and revolution, solstice and equinox, and seasonal variation of light and temperature. Students were encouraged to explore six different virtual models within the AR interface as part of the learning activity. They were also told to ask and answer questions to and from the instructor during the exercise.

The most physically active students were more successful in achieving a substantial change in understanding

In general, the students who were less physically active also asked fewer questions and initiated fewer exploratory movements of the virtual models. These same students made a less substantial change in their understanding from pre- to post-assessment than those who were more physically active. From the analysis, notes were taken with regard to the students who were noticeably less physically active during their instructional exercises, and then compared these notes with the student performances on the pre- and post-tests. Similarly, notes were written about the students who were very physically active during their instructional exercise, and compared these notes with their pre- and post-tests. This comparison confirms what even the most casual observer could identify, that the most physically active students were more successful in achieving a change in their understanding of Earth-Sun relationships.

Along this line of reasoning, Winn and Windschitl (2002) have noted that an important factor in which students differentiated learning strategies was
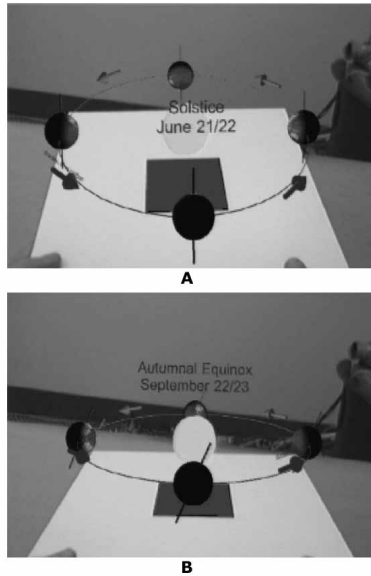
FIGURE 11
Comparing which part of the Earth is receiving light during equinox and solstice. In (A) the Earths on the left and right are equinoxes and the lit portions of the Earth are in-line with the axes. Rotating the card, such as pictured in (B), the viewer can see how the northern and southern hemispheres are unequally lit during solstices (Shelton, 2003).

the frequency of their use of tools in their virtual environment. In that case, the set of tools offered in the virtual problem solving environment consisted a set of built-in utilities. In the case of this AR exercise, the students used other kinds of tools afforded to them through the interface in order to complete their tasks. They used their normal investigative techniques of visual examination. Students turned objects at different angles to investigate the spatial relationships of the objects. They brought the objects closer to them to see greater detail. They used visual and spatial cues to determine light, position, and temporal arrangement of rotation/revolution, solstice/equinox, and seasonal variation of light. The back-and-forth nature of the "90 degree method" (see Figure 11) also qualifies as a learning strategy for those that used it repeatedly. After all, the "tool" given to the students in this situation consisted of virtual 3D objects and the physical nature of the interface itself. They had to use their powers of inspection through physical activity to find new information.

Most students who achieved larger changes in understanding moved the

card in a consistent cycle of "move, examine, and move" again. Some students preferred to leave the card at a particular perspective when trying to formulate answers to synthesis questions during times of investigator-student discussion. Questions of this nature, such as asking them about the length of days at certain times of year or how the temperature might be affected given the relative position of the Earth to the Sun, had a tendency to slow the interaction between student and interface. In these incidents, students were not focusing on any point on the model, instead, thinking about their own personal experiences to answer the question. But the majority of successful students exhibited more interaction with the virtual objects than those that did not.

Changing visual perspectives provided a measure of invariance for specific 3D inferences

For most students, changing their visual perspectives through their physical activity was crucial in understanding the relationship of the Earth's axis at different points in its orbital path. The physical activity mostly took the form of moving the position and angle of the card which resulted in having virtual objects either up-closer or farther-away, further to the left or to the right of the viewer. But some head movements in relationship to the card also took place. This changing of visual perspective provided a pathway necessary for the understanding solstice/equinox and seasonal variation, as well as the consistent axis angle of the Earth.

A good example of this point is a student pseudonamed Audra encountering the model in Figure 7 for the first time. Audra did not exhibit a theoretically accurate understanding of Earth-Sun relationships during her pre-assessment interview. She suspected that we experienced two solstices (one in summer, one in winter) but did not know what they meant. She admitted that summer might be caused due to the Earth being "closer to the Sun" during summer, but then did not think that it was a reason "that made sense."

At one point Audra identified the Earth closest to her by reading the annotation above the Earth. Audra associated each of the Earths with a season from her own experiences, realizing December is her "mid-winter," June is her "mid-Summer," and so on. She also noticed each Earth is equally spaced around the Sun on its revolutionary path and there were four of them. This also helped her reach her conclusion that each Earth represents a point in time of the year, reinforcing her idea that each Earth represents a season. Audra paused and "zoomed in" on an equinox. She noticed the angle of the equinox axis from her point of view. Audra turned the card in a direction

from her left to her right, looking at the Earth objects in a way consistent with the chronological order of the seasons. She examined and compared the axis angle of each Earth object. Audra understood that the axis of each Earth remains tilted at all times as she changes her perspective through her movement of the model.

Audra worked-out the notion of the tilt of the Earth's axis as it revolves around the Sun due to her physical movement and inspection of the axis as she turned the model. Her explanation of the Earth's position as it revolves around the Sun closely matched an expert during her post-assessment interview. In addition, the consistent turning between these 90 degree positions helped show the consistency of the Earth's tilt during its revolutionary path around the Sun. Students used this technique to determine the different areas illuminated on each Earth, as well as determining the consistent angle of the Earth during revolution.

Another student, pseudonamed Sally used the changing of visual perspectives to help her understand the consistency of the Earth's axis as it travels around the Sun. Sally continued to examine the Earths from different perspectives by turning the card back and forth on a one-quarter turn. She used a strategy similar to Audra, turning the card for a few moments, pausing for a few moments, then moving the card again. She alternated her focus to different Earths as she moved the card in this fashion. Sally needed a considerable amount of guidance to understand the consistency of the Earth's axis as it revolves around the Sun. Yet, it was the change in her perspective as she physically manipulated the card that helped her realize the 3D position and orientation of each of the Earth objects. This was a crucial piece of knowledge for students to have in order to advance to more complex topics such as how the seasonal variation of light and temperature are related to the Earth's position around the Sun.

Student control over the content was helpful

Seeing what they wanted to see, when they wanted to see it, provided the active student with control over the instructional content. In other words, the control over the content not only appeared beneficial, but students remarked on how and why this was important for them. This kind of control is in stark contrast to traditional forms of instruction, normally taking the form of text and 2D diagrams. Even more technologically advanced forms such as video or flash animations (dynamic non-interactive visualizations) do not let the student have full control over their examination of the content. Interactive 2D virtual content can offer a limited amount of control over the inspection

of virtual objects through an interface that allows the changing of viewing perspectives through button control.

Other evidence that students took advantage of the ability to control what they wanted to see, when they wanted to see it, was that some students preferred a more stable image. They preferred to move certain models only slightly, if at all, keeping the models more stable for their careful analysis. Other students preferred to move the cards often, sacrificing some stability in the rendering of the image but gaining appreciation for all aspects of the objects from many angles. Perhaps these students wanted to explore the entire aspect of every model to make sure they were not "missing" anything. All students moved the models a bit, but as previously stated, the more successful students were more active in changing perspectives with the models

In their work with learning strategies in virtual environments, Winn and Windschitl also found that more active students were more successful. "The most striking feature of Richard's explorations was his use of tools and *views in combination with one another* in order to make sense of his experiences" (Winn & Windschitl, 2002, p. 10). In the AR exercise, students used physical movement in combination with visual focus as their means to explore the details of the models. They then used comparative strategies as they changed perspectives to understand important spatial relationships between objects. Students learned about the seasonal variation of light and temperature. During Seattle's winter, the Southern Hemisphere receives the most daylight, and vice-versa. The cyclical nature of how Earth receives Sunlight is due to its angle of rotation. Seasonal variation of light is due to the way the Earth moves relative to the Sun. Seasonal variation of temperature is in part due to the way the Earth moves relative to the Sun.

The augmented reality interface lends itself well to task-related learning because of the exclusive connectivity between short cycles of visual perceptual activity and physical movements. This provided each student with the advantages for physical action of virtual content in the "real" world. Interfaces like AR can help explain about how people use their dual visual and physically interactive nature to learn. This evidence, plus data gathered in other formal and informal studies to date, suggests that AR provides greater cognitive access to more complex visualizations than conventional desktop interfaces. Further work is needed to corroborate these findings.

6. DISCUSSION

AR, Teaching and Learning: Some Implications for Success

In addition to the findings previously outlined in this section, we are interested in placing the results within a more established framework for learning using virtual environments. For this purpose, we draw upon the factors outlined by Winn and Windschitl for some comparisons between learning in completely immersive environments and those found here for augmented reality (Winn & Windschitl, 2002; Winn, 2001).

Winn and Windschitl described a number of factors that began to build a framework for learning in virtual environments. These factors were based on their studies with students in completely immersive VR. First, they described the importance of students being active participants in their learning. The more successful students in their immersive virtual environment were ones that explored more of their environment, asked more questions of the instructor, and were involved in more task-related activity. Similarly, we described the most successful AR students as being ones who tended to maneuver the virtual objects in many ways and who changed perspectives often to find the "most advantageous" positions in finding the information they wanted. We described how their visuo-motor activity was linked to how they learned, updating their readout strategies to encode new information.

Winn and Windschitl prescribed building a learning environment to maintain a sensible balance between student autonomy and providing guidance. The students who participated in the AR exercise needed guidance to learn, despite being presented with a rich visual environment. Learning did not simply "happen," the students went through a process of building new knowledge by modifying their causal nets. It was an iterative process that we demarcated throughout their activity. Students were consistently involved in the process of reorganizing their knowledge in a way that incorporated new knowledge elements with established ones.

Winn and Windschitl described how students and their environments constitute complex, interacting systems. In their study, students actively explored an immersive world where they moved around through a large space, stopping in different areas to interact with virtual elements. The students interacted with the virtual elements to record specific numerical values associated with each "area" of the virtual space. In this way, they were given the means to accomplish tasks aimed at reaching an assigned goal. "To teach, you must perturb the environment to induce adaptation" (p. 16).

The idea of allowing the student to become an active participant by perturb-

ing their environment follows the authors' point in how successful students differed in their "systemacity" within the virtual environment (Winn & Windschitl, 2002). In effect, the students were offered a way to make a change in certain variables and check their solution. In the Earth-Sun AR study, "systemacity" was employed not with "tools" built inside the AR system but rather in the way students used the affordances that existed within the interface. The AR student participants were allowed to use systematic physical/visual movements resulting in changes of perspective for spatially-related items. This way, AR students worked to coordinate their readout strategies within their environment. Winn and Windschitl (2002) also described the kinds of coordination allowed within their environment, and how the coordination occurred, as being extremely important in the learning process in their research.

The attributes of virtual objects in completely immersive VR hold the same advantages of virtual objects in a mostly "real" environment. "In artificial environments, reified abstractions have equal phenomenological status with models of real objects" (Winn & Windschitl, 2002, p. 17). Students have been afforded the advantages of virtual content in varying degrees since VR experiments involving education were first conducted (e.g., Byrne, 1993; Dede, 1995; McLellan, 1996; Osberg, Winn, Rose, Hollander, & Hoffman, 1997; Winn, Hoffman, Hollander, Osberg, & Rose, 1997; Youngblut, 1998). The virtual content involved in this AR study also used reified attributes to enhance the learning experience beyond what could normally be represented in classroom instruction. Students in the Earth-Sun AR research investigated artificial lighting that represented Sunlight and observed dynamic movements controlled by "unseen" elements that represented gravitational and celestial forces.

Winn and Windschitl also discuss the importance of studying the students and environment as a single system. Learning, as a social activity, does not exist in a vacuum. The collaboration and coordination between instructor, student, and technological interface consisted as a system, and therefore needed to be studied as one. With AR, to explore "how" students learned it was essential to view the system as a whole, and study it in an appropriate manner. Basing conclusions on the answers to the pre- and post-tests would not have told the whole story. The way students interacted with the interface and the instructor provided insight and evidence into which elements of the system had certain effects on the teaching/learning process. It is these elements which might be taken further into new research studies that focus in more detail on the design of the technology or the instruction, or may focus on developing theory for why the system worked as it did.

Future Research

The future for augmented reality as a visualization technology looks bright, as shown by the interest being generated in business and communications industrial circles and discussed in popular periodicals. But many questions still linger about its use in education. Is this a million-dollar solution to a ten-cent problem? Are 1st person manipulative AR instructional systems more effective than traditional methods to the degree that it should be implemented for learning about dynamic spatial relationships? Or are our visualization techniques satisfactory?

A research study that makes a direct comparison between the instructional design presented here, with AR, and more traditional techniques is needed to determine if students who used AR outperformed the students from the traditional group. The results could indicate the importance of using AR for learning dynamic spatial relationships. Is the implementation cost-effective? Using a one-on-one technique such as that employed here is costly. However, the same AR system could be used for different projects within the same discipline at different times. An example would be having the same system used to teach Earth-Sun relationships, moon phases and tide cycles, eclipses, solar system configurations, galaxy distributions, and so on.

The same system could be used across disciplines, such as astronomy, geography, and microbiology. We suggest that a follow-up study should use AR in a variety of topics that involve 3D dynamic relationships, such as learning about molecular interactions or geographical land formations.

Some research exists, and more is forthcoming, that studies the use of AR in collaborative settings (e.g., Fjeld et al., 2002; Kiyokawa et al., 2002; Takemura & Ohta, 2002). There is also interest in using AR for mobile settings, such as using it for navigation and for information within a museum (e.g., Grafe, Wortman, & Westphal, 2002; Lee, You, & Neumann, 2002; Wagner, 2002). Most of the current research is aimed at the development of the systems themselves, rather than empirical work of how the systems have been used for educational purposes. However, once the systems are implemented, research concerning how students use AR in collaborative learning situations and as mobile technologies could help inform classroom use as well.

AR also has the potential for expansion into multiple-user settings, such as multiple student classrooms and auditoriums. Little has been researched on these topics, a likely reason being the currently prohibitive costs of implementing AR on a large scale. However, as more interest is generated, additional software and hardware suppliers will enter the marketplace and a

study involving large-scale use (large settings, multiple participants) will be more feasible.

There is potential for studying aspects of virtual objects and environments that challenge the learning theory in this area. The design of the augmented reality system is critical to the success or failure of the Earth-Sun relationship instructional exercise. As previously noted, examining the use of visualization tools for educational purposes in context helps illuminate the way students understand instructional content. A research study aimed at the careful examination of the design factors in the instructional exercise will help in assessing the design of the interface for learning. The analysis would also suggest what kinds of content are supported or appropriate with the augmented reality system. We expect these design factors to include elements of the interface that define cues for the human visual system. "Presence," another design factor, could be defined and assessed using descriptors such as guidance, feedback, and levels of abstraction. It could then be more carefully compared with learning in completely immersive virtual environments (Winn, 2002; Winn, Windschitl, Fruland, & Lee, 2002).

In terms of virtual content, research needs to look at the design of the visual representations (3D objects) in aspects of movement, color, and size. Research is needed to examine the limitations and assets of the design choices that affect the students' understandings of the representations. In the process of this kind of research, other design factors are likely to be discovered as well.

Implications of AR for spatial knowledge acquisition

We have introduced augmented reality as representing interesting potential for spatial visualization due to its unique combination of viewing, manipulation and interaction characteristics. Much of cognitive theory predates the development of interfaces with such characteristics, so there is a need for existing theory to be integrated in order to provide a robust theoretical foundation as a basis for understanding their use. Drawing examples from spatial cognition, animate vision and learning theory, we proposed a set of linked theories to engage the use of these interfaces.

We expect continued experimentation in order to validate the hypothesis that AR holds advantages over other interfaces for certain tasks in which vision and task-related movement are critical. Examples of these tasks include geographic landmark recognition, velocity estimation and prediction, and the comprehension of physical laws.

Initial findings suggest that AR interfaces provide an as-yet unexplored

cognitive link between visualization and user. In addition to visual and sensorimotor feedback, the interface is set in a familiar spatial and visual context. This unique combination of affordances may provide a more direct cognitive relationship between users and spatial visualizations than conventional interfaces. Furthermore, this relationship may mediate the transformation of visualization content into spatial knowledge and is facilitated by visuo-motor factors affecting user interactions and the interface. We are engaged in empirical validation of these hypotheses, as a step towards informing future interface design, developing commensurate cognitive theory, and advancing spatial visualization practice. From informal and formal evidence gathered thus far, the intuitive nature of interacting with spatial information provided by AR may have the potential to provide more widespread intellectual access to spatial visualizations across novice and expert groups.

We suggest there are opportunities in which to take advantage of the AR interface in accordance with animate vision theory. Due to exclusive connectivity between short cycles of visual perceptual activity and physical movements, the AR interface lends itself well to task-related learning or study using space and the spatial derivatives of velocity and acceleration. Following animate vision theory, this interface draws its strength upon the properties that are not used for verbal or gestural statements and is therefore immune from certain visual illusions. Perhaps it is most advantageously used for action in the world, and physical processes that involve action by the participant.

Research in AR has implications for further integration of cognitive psychology theory with visual theory for artificial systems. The potential for combining disciplines may continue to generate ideas concerning how people see and make decisions based on the physical and spatial nature of their environment. Disciplines such as industrial design, architecture, and medicine may find a new way of teaching students about complex spatial phenomena within their discipline. The research may also point toward improvements or changes in the design of the augmented reality interface in order to better meet the needs of students. In the age of ever-expanding technological advances, finding new ways for people to experience and construct knowledge should remain at the forefront.

NOTES

Background

A large portion of the section on AR and education content was adapted from a paper in New Horizons for Learning (Shelton, 2002). The authors have been collaborating on the use of AR for education since March 2002, as researchers for the University of Washington's Program for Educational Transformation Through Technology (PETTT).

Acknowledgments

REFERENCES

Aloimonos, Y. (1993). Introduction: Active vision revisited. In Y. Aloimonos (Ed.), *Active Perception* (pp. 1-18). Hillsdale, NJ: Lawrence Erlbaum Associates.

Azuma, R. T. (1997). A survey of augmented reality. *Presence, 6*(4), 355-385.

Azuma, R. T., Baillot, Y., Behringer, R., Feiner, S., Julier, S., & MacIntyre, B. (2001). Recent advances in augmented reality. *IEEE Computers and Graphics*(November/December 2001), 34-47.

Ballard, D. (1991). Animate vision. *Artificial Intelligence, 48*, 57-86.

Barab, S. A., Hay, K. E., Barnett, M., & Keating, T. (2000). Virtual Solar System project: Building understanding through model building. *Journal of Research in Science Teaching, 37*(7), 719-756.

Billinghurst, M., Kato, H., & Poupyrev, I. (2001). The Magic Book — Moving seamlessly between reality and virtuality. *IEEE Computers, Graphics and Applications, 21*(3), 2-4.

Byrne, C. M. (1993). *Virtual Reality in Education* (HITLab Technical Report R-93-6). Seattle: University of Washington.

Byrne, C. M. (1996). *Water on Tap: The Use of Virtual Reality as an Educational Tool*. Unpublished Dissertation, University of Washington, Seattle.

Cheok, A. D., Weihua, W., Yang, X., Prince, S., Wan, F. S., Billinghurst, M., & Kato, H. (2002). *Interactive theater experience in embodied + wearable mixed reality space*. Paper presented at the International Symposium on Mixed and Augmented Reality (ISMAR '02), Darmstadt, Germany.

Clark, A. (1997). *Being There: Putting Brain, Body, and World Together Again*. Cambridge, MA: The MIT Press.

Dede, C. (1995). The evolution of constructivist learning environments: Immersion in distributed, virtual worlds. *Educational Technology, 35*(5), 46-52.

Dienes, Z., & Perner, J. (1999). A theory of implicit and explicit knowledge. *Behavioral and Brain Sciences, 22*, 735-808.

Ditlea, S. (2003, September 20, 2002). Augmented reality. *Popular Science*.

Drascic, D., & Milgram, P. (1996). *Perceptual issues in augmented reality*. Paper presented at

the SPIE, Stereoscopic Displays VII and Virtual Systems III, Bellingham, WA.

Feiner, S. K. (2002, April 2, 2002). Augmented reality: A new way of seeing. *Scientific American*.

Fjeld, M., Schar, S. G., Signorello, D., & Krueger, H. (2002, September 30, 2002). *Alternative tools for tangible interaction: A usability evaluation.* Paper presented at the IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR), Darmstadt, Germany.

Golledge, R. G. (1991). Cognition of physical and built environments. In T. Garling & G. W. Evans (Eds.), *Environment, Cognition and Action — An Integrated Approach* (pp. 35-62). New York: Oxford University Press.

Grafe, M., Wortman, R., & Westphal, H. (2002, September 29, 2002). *AR-based interactive exploration of a museum exhibit.* Paper presented at the First IEEE International Augmented Reality Toolkit Workshop, Darmstadt, Germany.

Haken, H. (1991). *Synergetic Computers and Cognition*. Berlin: Springer.

Hedley, N. R. (2001, March, 2001). *Exploring the dimensions of spatial mental models formed in augmented and virtual environments.* Paper presented at the Association of American Geographers Annual Meeting, New York, NY.

Hedley, N. R. (2003a). *Empirical evidence for advanced geographic visualization interface use.* Paper presented at the 2003 International Cartographic Association's International Cartographic Congress (ICC 2003), Los Alamitos, CA.

Hedley, N. R. (2003b). *Three-dimensional geographic visualization and spatial mental models.* Unpublished Doctoral Dissertation, University of Washington, Seattle, WA.

Hedley, N. R., & Shelton, B. E. (2001). *Spatial knowledge acquisition and augmented reality: Exploring theory and practice in empirical settings* (Report for the Program for Educational Transformation Through Technology). Seattle: University of Washington.

HITLab (2002). *3D model enhanced with augmented reality for learning biology and chemistry*: personal communication with the Human Interface Technology Laboratory and the TSRI Molecular Graphics Laboratory.

Kiyokawa, K., Billinghurst, M., Hayes, S. E., Gupta, A., Sannohe, Y., & Kato, H. (2002, September 30, 2002). *Communication behaviors of co-located users in collaborative AR interfaces.* Paper presented at the IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR), Darmstadt, Germany.

Lee, W. L., You, S., & Neumann, U. (2002, September 30, 2002). *Tracking with omni-directional vision for outdoor AR systems.* Paper presented at the IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR), Darmstadt, Germany.

Lowe, R. (1993). Constructing a mental representation from an abstract technical diagram. *Learning and Instruction, 3*, 157-179.

Mackay, D. (1967). Ways of looking at perception. In W. Wathen-Dunn (Ed.), *Models for the Perception of Speech and Visual Form* : MIT Press.

Mark, D. M. (1992). *Spatial metaphors for human-computer interaction.* Paper presented at the Fifth International Symposium on Spatial Data Handling.

Mark, D. M. (1993). Human spatial cognition. In D. Medychkyj-Scott & H. M. Hearnshaw (Eds.), *Human Factors in Geographic Information Systems* . London: Bellhaven Press.

Mark, D. M., & Freundschuh, S. M. (1995). Spatial concepts and cognitive models for geographic information use. In T. L. Nyerges, D. M. Mark, R. Laurini, & M. Egenhofer (Eds.), *Cognitive Aspects of Human-Computer Interaction for Geographic Information Systems* . Dordrecht: Kluwer Academic Publishers.

Marr, D. (1982). Chapter 1: The philosophy and the approach, *Vision* : Freeman.

McLellan, H. (1996). Virtual realities. In D. Jonassen (Ed.), *Handbook of Research for*

Educational Communications and Technology . New York: Simon & Shuster Macmillan.

Milgram, P. (1994). Augmented reality: A class of displays on the reality-virtuality continuum. *SPIE Volume 2351: Telemanipulator and Telepresence Technologies*.

Neisser, U. (1976). *Cognition and Reality: Principles and Implications of Cognitive Psychology*. San Francisco: W. H. Freeman and Company.

Neumann, U., & Majoros, A. (1998). *Cognitive, performance, and systems issues for augmented reality applications in manufacturing and maintenance.* Paper presented at the IEEE Virtual Reality Annual International Symposium (VRAIS 98), Los Alamitos, CA.

O'Regan, J. (1992). Solving the "real" mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology, 46*, 461-488.

Osberg, K. (1993). *Virtual reality and education: A look at both sides of the sword* (HITLab Technical Report R-93-7). Seattle: University of Washington.

Osberg, K., Winn, W., Rose, H., Hollander, A., & Hoffman, H. (1997). *The Effect of Having Grade Seven Students Construct Virtual Environments on their Comprehension of Science.* Paper presented at the Annual Meeting of the American Educational Research Association, Chicago.

Portugali, J. (1996). Inter-representation networks and cognitive maps. In J. Portugali (Ed.), *The Construction of Cognitive Maps* (pp. 11-43). Dordrecht: Kluwer Academic Publishers.

Rayner, K., Well, A., & Pollarsek, A. (1980). Asymmetry of the effective visual field in reading. *Perceptual Psychophysics, 27*, 537-544.

Rolland, J. P., & Fuchs, H. (2000). Optical versus video see-through head-mounted sidplays in medical visualization. *Presence: Teleoperators and Virtual Environments, 9*(3), 287-309.

Shelton, B. E. (2002). Augmented reality and education: Current projects and the potential for classroom learning. *New Horizons for Learning, 9*(1).

Shelton, B. E. (2003). *How augmented reality helps students learn dynamic spatial relationships.* Unpublished Doctoral Dissertation, University of Washington, Seattle.

Shelton, B. E., & Hedley, N. R. (2001). *An enhanced theoretical basis for affordance-derived interface design* (Report for the Program for Educational Transformation Through Technology). Seattle: University of Washington.

Shelton, B. E., & Hedley, N. R. (2002, September 29, 2002). *Using augmented reality for teaching Earth-Sun relationships to undergraduate geography students.* Paper presented at the First IEEE International Augmented Reality Toolkit Workshop, Darmstadt, Germany.

Shelton, B. E., Humble, J., & Matson, J. (1996, October 1996). *Empirical Study of Effects on Perception of Computer Graphic Reenactments.* Paper presented at the Contemporary Issues of Employment Practices and Principles, San Diego, CA.

Takemura, M., & Ohta, Y. (2002, September 30, 2002). *Diminishing head-mounted display for shared mixed reality.* Paper presented at the IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR), Darmstadt, Germany.

Wagner, M. (2002, September 29, 2002). *Building wide-area applications with the AR Toolkit.* Paper presented at the First IEEE International Augmented Reality Toolkit Workshop, Darmstadt, Germany.

Winn, W. (1995). Virtual reality in the classroom: The virtual reality roving vehicle project. *Technological Horizons in Education, 23*(5), 70-74.

Winn, W. (2002). Learning in artificial environments: Embodiment, embeddedness and dynamic adaption. *Technology, Instruction, Cognition and Learning, 1*, 28.

Winn, W., Hoffman, H., Hollander, A., Osberg, K., & Rose, H. (1997). *The Effect of Student Construction of Virtual Environments on the Performance of High- and Low-Ability Students.* Paper presented at the Annual Meeting of the American Educational Research Association, Chicago.

Winn, W., & Snyder, D. (1996). Chapter 5: Cognitive perspectives in psychology. In D. H. Jonassen (Ed.), *Handbook of Research for Educational Communications and Technology* (pp. 1107-1136). New York: Simon and Schuster Macmillan.

Winn, W., & Windschitl, M. (2002, April 2002). *Strategies used by university students to learn aspects of physical oceanography in a virtual environment.* Paper presented at the American Educational Research Association, New Orleans, LA.

Winn, W., Windschitl, M., Fruland, R., & Lee, Y. (2002, October 23-26, 2002). *When does immersion in a virtual environment help students construct understanding?* Paper presented at the Internactional Conference on the Learning Sciences (ICLS), Seattle, WA.

Winn, W. D. (1994). Contributions of perceptual and cognitive processes to the comprehension of graphics. In W. Schnotz & R. W. Kulhavy (Eds.), *Comprehension of Graphics* . Amsterdam: North-Holland.

Winn, W. D. (2001). Current trends in educational technology research: The study of learning environments. *Educational Psychology Review*.

Youngblut, C. (1998). *Educational uses of virtual reality technology* (IDA Document D-2128). Alexandria, VA: Institute for Defense Analyses.