2017

Getting the Upper Hand: Natural Gesture Interfaces Improve Instructional Efficiency on a Conceptual Computer Lesson

Shannon Bailey
University of Central Florida

GETTING THE UPPER HAND: NATURAL GESTURE INTERACTIONS IMPROVE INSTRUCTIONAL EFFICIENCY ON A CONCEPTUAL COMPUTER LESSON

by

SHANNON KYLE TEDDER BAILEY
B.S. University of Central Florida, 2012
M.A. University of Central Florida, 2016

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Psychology
in the College of Sciences
at the University of Central Florida
Orlando, Florida

Fall Term
2017

Major Professor: Valerie K. Sims

ABSTRACT

As gesture-based interactions with computer interfaces become more technologically feasible for educational and training systems, it is important to consider what interactions are best for the learner. Computer interactions should not interfere with learning nor increase the mental effort of completing the lesson. The purpose of the current set of studies was to determine whether natural gesture-based interactions, or instruction of those gestures, help the learner in a computer lesson by increasing learning and reducing mental effort. First, two studies were conducted to determine what gestures were considered natural by participants. Then, those gestures were implemented in an experiment to compare type of gesture and type of gesture instruction on learning conceptual information from a computer lesson. The goal of these studies was to determine the instructional efficiency – that is, the extent of learning taking into account the amount of mental effort – of implementing gesture-based interactions in a conceptual computer lesson.

To test whether the type of gesture interaction affects conceptual learning in a computer lesson, the gesture-based interactions were either naturally- or arbitrarily-mapped to the learning material on the fundamentals of optics. The optics lesson presented conceptual information about reflection and refraction, and participants used the gesture-based interactions during the lesson to manipulate on-screen lenses and mirrors in a beam of light. The beam of light refracted/reflected at the angle corresponding with type of lens/mirror. The natural gesture-based interactions were those that mimicked the physical movement used to manipulate the lenses and mirrors in the optics lesson, while the arbitrary gestures were those that did not match the movement of the lens or mirror being manipulated. The natural gestures implemented in the

computer lesson were determined from Study 1, in which participants performed gestures they considered natural for a set of actions, and rated in Study 2 as most closely resembling the physical interaction they represent. The arbitrary gestures were rated by participants as most arbitrary for each computer action in Study 2. To test whether the effect of novel gesture-based interactions depends on how they are taught, the way the gestures were instructed was varied in the main experiment by using either video- or text-based tutorials.

Results of the experiment support that natural gesture-based interactions were better for learning than arbitrary gestures, and instruction of the gestures largely did not affect learning and amount of mental effort felt during the task. To further investigate the factors affecting instructional efficiency in using gesture-based interactions for a computer lesson, individual differences of the learner were taken into account. Results indicated that the instructional efficiency of the gestures and their instruction depended on an individual's spatial ability, such that arbitrary gesture interactions taught with a text-based tutorial were particularly inefficient for those with lower spatial ability. These findings are explained in the context of Embodied Cognition and Cognitive Load Theory, and guidelines are provided for instructional design of computer lessons using natural user interfaces.

The theoretical frameworks of Embodied Cognition and Cognitive Load Theory were used to explain why gesture-based interactions and their instructions impacted the instructional efficiency of these factors in a computer lesson. Gesture-based interactions that are natural (i.e., mimic the physical interaction by corresponding to the learning material) were more instructionally efficient than arbitrary gestures because natural gestures may help schema development of conceptual information through physical enactment of the learning material.

Furthermore, natural gestures resulted in lower cognitive load than arbitrary gestures, because arbitrary gestures that do not match the learning material may increase the working memory processing not associated with the learning material during the lesson. Additionally, the way in which the gesture-based interactions were taught was varied by either instructing the gestures with video- or text-based tutorials, and it was hypothesized that video-based tutorials would be a better way to instruct gesture-based interactions because the videos may help the learner to visualize the interactions and create a more easily recalled sensorimotor representation for the gestures; however, this hypothesis was not supported and there was not strong evidence that video-based tutorials were more instructionally efficient than text-based instructions. The results of the current set of studies can be applied to educational and training systems that incorporate a gesture-based interface. The finding that more natural gestures are better for learning efficiency, cognitive load, and a variety of usability factors should encourage instructional designers and researchers to keep the user in mind when developing gesture-based interactions.

ACKNOWLEDGMENTS

I would like to thank my dissertation committee chair, Dr. Valerie Sims, and my committee members, Drs. Florian Jentsch, Clint Bowers, and Cheryl Johnson, for their valuable input. I am grateful for their expertise that led to the development of the theory, methodology, and analyses of this dissertation. I would also like to acknowledge the immense effort of programmers William McKenna and Matt Proetsch, who worked many underpaid hours to develop the computer lesson. Without their dedication, I would not have been able to take my ideas from doodles on paper to a fully-functioning testbed, and this dissertation would not exist. You are both worth way more than the doughnuts you were paid in. I want to thank my friends and family who have supported me through it all: Aislinn, Mom, Dad, Nina, Luke, grandparents, Lexi, Lindsay, Mindy, Nick, Joanna, ACAT lab, STRIKE lab, and so many others. Finally, I would not be where I am today without the teachers who led me here. Thank you for dedicating your lives to public education and the lives of others. Your encouragement truly made my life possible.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE: INTRODUCTION

As motion tracking technology becomes more accurate and widely available, it is feasible to implement gesture-based interactions in systems for education and training. Before gesture-based interactions should be included in educational systems, it is important to understand how such interfaces affect learning to avoid implementing interactions that may negatively affect the learner. For example, arbitrary gesture-based interactions that do not match the learning material may hinder learning because interacting with the lesson is an additional mental burden on the learner that is not relevant to the lesson; however, gesture-based interactions in a computer lesson that match the learning material, or natural gestures, may be easier to use and foster stronger memories for the learning material than arbitrary gesture-based interactions that do not match the learning material. Alternatively, the way in which the gestures are instructed may influence the feelings of naturalness for the interactions and impact learning more than how much the interaction matches the learning material. The current set of studies investigates whether natural gesture-based computer interactions that match the learning material are a more beneficial instructional technique than arbitrary gesturing, or whether the instruction of the interaction matters more for learning.

Gesture-based Interactions

Controls that are more intuitively mapped may facilitate ease of interacting with a computer system (Norman, 2002). Recent advances in technology may make interacting with a computer interface more intuitive, such as gesture-based commands that more closely mimic the physical actions they represent (Dodds, Mohler, & Bulthoff, 2011; Singer & Goldin-Meadow, 2005). Imagine a surgeon standing by the bedside of a patient under anesthesia, viewing a monitor displaying a 3D image of the patient's internal organs. She moves her gloved hand in front of the screen in a rotating motion, and the image on the screen turns in response. The surgeon gestures again, this time the image zooms in for a better view of the area in which she will be operating. A motion tracker captures the surgeon's gestures, and the gestural commands manipulate the image displayed on the monitor in response. The surgeon is able to quickly change the image while keeping her hands in a sterile zone, without breaking the mental flow of surgery. Motioning her hand to move an image on the screen more closely maps onto the physical process of moving an object than the traditional computer interaction of using a mouse and pointer.

As motion tracking technology becomes more widely available, natural gesture-based interactions may be implemented into educational and training computer systems, which could have instructional benefits over less intuitive computer interactions. In the case of a conceptual computer lesson, controls that are more intuitively mapped may facilitate ease of interacting because natural gestures might reduce the cognitive load, or amount of information being processed, of the learner (Goldin-Meadow, Nusbaum, Kelley, & Wagner, 2001; Hamblin, 2005; Wagner, Nusbaum, & Goldin-Meadow, 2004), and build stronger mental representations of the conceptual material by enacting, or physically performing, the gestures (Engelkamp & Zimmer, 1997; Schwartz & Plass, 2014). Research has shown that computer interactions that represent a physical action (e.g., dragging the mouse to move an icon) are easier to remember than arbitrary interactions that do not correspond with actions in the real world or the learning material (e.g., clicking the mouse to move an icon; Schwartz & Plass, 2014). The finding that enactment helps memory is referred to as the *enactment effect* and is part of a multi-system framework that combines conceptual and sensory information during encoding to produce stronger memories (Engelkamp & Jahn, 2003). It follows that using gestures to interact with a computer lesson may be easier and more memorable if the computer interactions more closely match the physical actions they represent.

Instructing Gestural Interactions

Alternatively, it could be that the ease of gesture-based interaction is dependent on how the gestures are instructed. Traditionally, using a computer involves an interface in which a user is taught to move a proximal device (i.e., a mouse) to control a distal object on a screen (i.e., pointer) using learned mechanisms (e.g., double clicking the mouse to select, using the scroll wheel to zoom, etc.). Actions such as double clicking to select or scrolling to zoom are only arbitrarily mapped to the on-screen actions they represent, but interacting with a mouse may seem second nature once the process is instructed. Instructions may help users overcome "conceptual difficulties" of learning novel gesture-based interactions and make a gesture seem more intuitive (p. 251, Schurmann, Binder, Janzarik, & Vogt, 2015). Just as people learned to use a mouse to perform computer tasks, perhaps the naturalness of gesture-based computer interactions depends on how well those interactions are instructed.

The question of how gestures and their instruction may support conceptual learning in a computer environment can be explained though the theoretical frameworks of Embodied Cognition (Barsalou, 2008; Wilson, 2002) and Cognitive Load Theory (CLT; Sweller, van Merriënboer, & Paas, 1998; Sweller, 2010). The term *Embodied Cognition* encompasses many theories that can be summarized by the tenant that one's physical interactions with the world can shape thinking. For example, when actions are physically performed or observed, they can activate the motor system, which serves to create stronger memories and develop schemas for those actions (Barsalou, 2008; Engelkamp & Jahn, 2003; Hostetter & Alibali, 2008). Like theories of embodiment, the CLT framework may also explain why these instructional techniques affect learning, by contextualizing different aspects of information processing. CLT is based on the idea that new information is processed by working memory to develop representations of that information for storage in long term memory, but the capacity in working memory for new information is limited (Sweller, van Merriënboer, & Paas, 1998). The capacity in working memory is filled by different kinds of information processing, or cognitive load. A goal of instructional design, therefore, is to reduce the cognitive load that is not useful for developing mental representations (i.e., learning). CLT may help explain why gesture-based interactions, along with how they are instructed, support or hinder learning.

Deficiencies in Studies

Research is needed on how gesture-based interactions impact learning conceptual information in human-computer systems. Previous research on computer interface interactions have focused on the amount of interaction a user has with the system, finding that more interactivity leads to better task performance and recall (see Betrancourt, 2005 for a review of interactivity in multimedia systems). For example, research has determined that memory for computer actions is better when physically moving a mouse to control the action than simply viewing an action completed on-screen (Schwartz & Plass, 2014). In addition to the amount of interaction, research has compared types of computer interaction, including the level of natural mapping between the interaction and the real-world action represented (Norman, 2002). The more natural and less arbitrary an interaction, the more likely the interaction is to be recalled later (Schwartz & Plass, 2014). At the same time, instructional design research emphasizes the importance of appropriate instruction for a task (Mayer & Moreno, 2010), such as choosing the instruction's type of media (e.g., text, picture, etc.; Zacks & Tversky, 2003) or modality (e.g., physical, verbal, etc.; Nilsson, Cohen & Nyberg, 1989) to match the learning material.

What is lacking from these previous studies is that there may be overlap between the type of interaction and how those interactions are instructed. For example, if people find a gesture command to be intuitive, is it because they have been well-instructed on how to perform that gesture (e.g., double clicking a mouse) and would they find any gesture intuitive with enough experience or instruction? In contrast, perhaps no amount of instruction can overcome a system of interaction that is so unintuitive that it does not make sense. The type of interaction may impact learning outcomes more or less depending on their instruction, or combined implementation of these factors may interact to produce mitigating or strengthening effects on conceptual learning. Therefore, research should investigate the interplay of interaction and instruction to differentiate their effects.

Purpose Statement

The current study investigated whether gesture-based interactions, and how the gestures were instructed, impacted learning conceptual information from a computer lesson. The computer lesson involved learning conceptual information about optics, such as how light interacts with mirrors and lenses. Participants used gestural interactions to complete the computer lesson that were either naturally mapped to the learning material, corresponding to the movement of the on-screen mirrors and lenses, or were arbitrary gestures irrelevant to the conceptual material. The gesture-based interactions were instructed using either video- or text-based tutorials to determine whether instruction of the interaction affected the understanding of the gestures and their effect on the computer lesson.

Significance for Application and Theory

These important questions about how gesture-based interactions affect conceptual learning should be answered as technology moves forward in educational and training systems.

Research is needed on how more intuitive interactions, such as gesturing, impact human-computer systems and whether those interactions are greatly affected by their instruction. The results will inform the extent to which computer interactions affect learning depending on the nature of the interaction (e.g., natural or arbitrary gesture; Dodds, Mohler, & Bulthoff, 2011; Singer & Goldin-Meadow, 2005) or on how the interaction is instructed (e.g., video versus text instruction; Engelkamp & Zimmer, 1997). By testing these factors in one experiment, the combined effects of computer interactions and their instructions can be compared to determine their joint instructional efficiency, and practical guidelines for application in future systems can follow. In addition to the usefulness of these results for designing educational computer systems, this research has important implications for theories of Embodied Cognition and CLT. Results will inform theories of Embodied Cognition by providing evidence of whether enacting the learning material through naturally-mapped computer interactions is better for learning conceptual information than arbitrarily-mapped gestures, and whether viewing the gestures in video-based instructions is more effective than text-based instructions. The findings will also suggest how well CLT explains the instructional efficiency of these interactions and their instructions by indicating the mental effort associated with each technique.

CHAPTER TWO: LITERATURE REVIEW

Theoretical Frameworks

The goal of the current set of studies is to provide evidence for appropriate gesture-based interactions and instruction for educational computer systems that is grounded in cognitive science. The theoretical frameworks from which this research was developed are Embodied Cognition and Cognitive Load Theory. First, an overview of each is described in general terms, then specific theories or tenants under these frameworks are presented in context of gesture interactions and instruction later in the chapter.

Embodied Cognition

Embodied Cognition theories, also referred to as *Grounded Cognition* or *Situated Cognition*, focus on how thinking is shaped by physical interactions with one's environment, stating that mental representations of information are not merely a series of verbal proposition statements (Barsalou, 2008; Garbinia & Adenzato, 2004; Wilson, 2002). Concepts are built, and thus dependent, on the sensory state in which the information was received and are not completely abstracted from that modality (Barsalou, 2008). Bodily-specific interactions on objects cannot be separated from cognition about those objects. For example, if a water bottle sits on a table, a person who is tall with a longer arm reach will perceive the water bottle as closer than a person who is shorter with a smaller reach (Longo & Lourenco, 2007). According to embodiment theories, this difference in spatial perception is due to the specific physical interactions the tall and short people have had with the world that created different representations of space.

One theory within the embodied cognition framework is simulation theory, which argues that mental representations exist in a neural sensorimotor system that correlates action and perception in modality-specific states (Barsalou, Barbey, Simmons, & Santos, 2005). To clarify, modal states are activated during cognition by sensory information, such as hearing a sound (i.e., auditory modality) or grasping an object (i.e., motor modality), creating a mental simulation of the original sensory experience in which one can mentally imagine hearing that sound or grasping an object. During perception and cognition of sensory information in which the body senses modality-specific information and contextualizes it, concepts can be encoded into memory for later activation, or simulation, in the sensory modality of the original stimulus. This reenactment may be partial or distorted (Barsalou et al., 2005). There is behavioral evidence for the simulation theory in research on the enactment effect. The *enactment effect* is the finding that physically-performed actions are more accurately recalled and retained longer than when the information was acquired in another modality (Engelkamp & Jahn, 2003). Engelkamp and Jahn (2003) explain the enactment effect as a result of a combined conceptual and sensori-motor multisystem that is activated during modality-specific encoding and retrieval of action phrases causes a "regeneration" of motor information that was encoded physically. That is, physically performing the actions encoded that motor information as part of the concept for that action and then recalling that concept involves simulating (i.e., "regenerating") that motor information. The enactment effect is described in the context of natural gesture-based computer interactions later in this chapter.

Neural evidence also supports this reactivation principle of simulation theory. In multiple areas of the brain, individual neural cells and combined activation patterns indicate that stored information retains the form of the original stimulus and are interconnected with other modalities (Barsalou, 2008; Garbarini & Adenzato, 2004; Goodale & Humphrey, 1998; Pezzulo et al., 2011; Rizzolatti & Craighero, 2004). For example, Rizzolatti and Craighero (2004) explain that mirror neurons activate for a specific action, such as grasping. Mirror neurons can be activated in primates when observing someone else perform that action, thus the name "mirror" neurons. The embodiment theorists explain that this activation of an action concept (e.g., grasping) in that sensory modality when the action was not physically performed is evidence for simulation theory, because the action was being represented "as if" it were being performed. Researchers took this experiment a step further, by addressing if the activation was merely visual recognition of an action pattern or an actual mental simulation of the action. To remove the possibility of visual recognition, Kohler et al. (2002) explored whether mirror neurons were activated with other modalities. The study found that the same neuron was activated in conditions in which a monkey physically cracked a nut, visually observed another cracking a nut, or heard a nut being cracked without seeing it. In each case, the same neuron fired for the concept of "nut cracking" as if the action were happening, regardless of modality. The ability to mentally simulate is considered an important mental phenomenon, because simulation can create a better understanding of goals and actions of others, by acting "as if" the observed action is happening to one's self (Barsalou, 2008). Empirical evidence supports the simulation principle that representations are formed by modality-specific sensations and retain those modalities when activated for cognition.

Cognitive Load Theory

Sweller, van Merriënboer, and Paas (1998) argued that the goal of instructional designers is to create educational materials that facilitate schema development, and those schemas are constructed via working memory processing. Cognitive Load Theory (CLT) is a theoretical framework from which instructional materials can be designed that take into consideration a learner's ability to process information in working memory. Before describing the ways in which instructional materials can facilitate schema development, I first describe the cognitive mechanisms underlying schema development, or the process by which incoming information from a lesson is encoded into long term mental representations for that information.

The cognitive mechanism behind the processing of information into stored mental representations is described in Dual Coding Theory (DCT; Clark & Paivio, 1991). Clark and Paivio (1991) state that imagery and linguistic information are processed separately in nonverbal and verbal structures into long term mental representations. Within each structure, either verbal or nonverbal information is processed such that associative connections are made between stored representations in that structure. Additionally, referential connections can be made between the verbal and nonverbal mental representations. A mental representation is therefore stored with connections to other mental representations, so when a mental representation for certain information is activated, that activation can spread to other mental representations for information that are connected to that first activation. The strength of these associative connections within each structure and the referential connections between the structures determines the ease with which those mental representations are recalled because the representations become activated from connection to other representations. These cognitive mechanisms described in DCT are related to the previously discussed theories of Embodied Cognition in that information encoded in one modality (e.g., visual information) is stored in a mental representation for that information that can be later recalled when that modality is activated (e.g., seeing an action can activate the neural pattern for that mental representation).

DCT describes the verbal and nonverbal systems in which information is processed into long term mental representations, and other theories of working memory extend this theory by suggesting the processing of each system is limited in how much information can be processed at a time. Like DCT, Baddeley and Hitch's model of working memory (Baddeley, 2000; Baddeley & Hitch, 1974) suggested that verbal and nonverbal information are processed in separate systems, the visuospatial sketchpad and the phonological loop. They explained that processing too much information in either system may overload that system, but this limitation can be mitigated by offloading information from one system to another. For example, if a lesson involves teaching a concept by having the learner read textual information, that verbal information is processed in the phonological loop. If additional information is presented to the learner as narrated speech, this verbal information may also be processed in the phonological loop, overwhelming the verbal processing system with too much information at once. Alternatively, if the new information were presented as a picture (i.e., visual information) instead of as narrated speech (i.e., verbal information), neither the phonological loop nor the visuospatial sketchpad are overwhelmed with multiple pieces of information to process at once. As suggested by research described later in this chapter, gestures may facilitate learning by offloading information processing that would otherwise overwhelm the verbal system of working memory.

The mechanisms described by DCT and the working memory model by Baddeley and Hitch provide cognitive explanations for the instructional design strategies provided by CLT. As explained in DCT, working memory processes verbal and nonverbal information into mental representations for that information, while forming connections to other stored information. These mental representations can also be described as schemas for learning material presented in a lesson. Sweller and colleagues (1998) explained that there are three types of cognitive load that impact working memory and thus affect schema development: intrinsic load, extraneous load, and germane load. Intrinsic load is the mental processing associated with the learning material itself, with more difficult material creating more intrinsic load. The difficulty of the learning material is based on how many elements must be processed concurrently, or the element interactivity (Sweller, 2010), with most CLT theorists holding the assumption that intrinsic load cannot be altered (Moreno & Park, 2010). Extraneous cognitive load is the way the learning material is presented to the learner that imposes unnecessary mental processing not related to the learning material (Sweller et al., 1998), which creates more interacting elements to be processed that are not relevant to learning (Sweller, 2010). For example, if a computer lesson involves bells and whistles unrelated to the learning material, the learner's working memory may be taxed with understanding why the sounds are occurring, leaving less working memory capacity for processing the relevant information into a schema for the lesson. Whereas intrinsic load associated with the learning material cannot be changed, extraneous load can be reduced through appropriate instructional techniques. The last component of cognitive load is germane load, or the mental processing related to schema development. Sweller (2010) argues that available working memory should be used by the germane load associated with processing information into schemas, instead of irrelevant extraneous load.

The three types of cognitive load are additive, and their combined effort can exceed the capacity of working memory (Sweller et al. 1998). When there is too much information to be processed in working memory, that information may not be developed into a schema or mental representation in long term memory. If the combined cognitive load is too great to process the new information, learning does not occur; therefore, the goal of a lesson should be to manage the cognitive load with the most efficient instructional design. Instructional efficiency can be measured by how much is learned from a lesson in relation to how much mental effort was used (Paas & van Merriënboer, 1994). Sweller et al. (1998) explain that because intrinsic load is inherent to the learning material, instructional efficiency is created by reducing extraneous load and redirecting attention to the relevant material processed via germane load. Mayer and Moreno (2010) suggested ways in which extraneous load can be reduced in instructional design include adherence to the *coherence principle*, or removing extraneous material from a lesson (e.g., irrelevant sounds, seductive details), and the *signaling principle*, or directing the learner's attention to the learning material (e.g., highlighting, headings), among others. Gesture-based interactions used in a computer lesson may impact either of these principles by either drawing attention to the learning material that correspond to the natural gestures or, if the gestural interactions do not match the learning material, violating the coherence principle. The way in which gesture-based interactions and how they are instructed can impact instructional efficiency is explained using cognitive load theory throughout this thesis.

Additionally, individual differences of the learner may also affect the level of cognitive load felt during instruction depending on the instructional technique, such as the learner's spatial ability or prior knowledge. The individual differences that may affect learning of the conceptual information in the current experiment were included as potential confounds, including the learner's spatial ability and prior knowledge. The learner's spatial ability may affect the instructional efficiency of the lesson because the optics concepts in the lesson are inherently spatial (i.e., how light reflects/refracts at different angles depending on the rotation of various lenses/mirrors). Spatial ability has been linked to the ability to understand mental simulation tasks that are similar to the current optics task, such that those with higher spatial ability are better able to make inferences about motion than lower spatial individuals, reducing the working memory associated with processing that information for people with high spatial ability (Hegarty & Sims, 1994). Therefore, there may be differences in the instructional efficiency depending on spatial ability in the current study, because high spatial learners may understand the mental animation of the lesson better than people with low spatial ability, corresponding with less cognitive load for those with high spatial ability. On the other hand, natural gestures that physically enact the motion of the optics task may assist the lower spatial individuals to understand the motion involved in the optics lesson, such that those with low spatial ability experience less cognitive load with natural gestures than arbitrary gestures. Another individual difference that could affect cognitive load is the learner's prior knowledge of optics before the computer lesson. Prior knowledge could help the learner in the current lesson because those with higher optics knowledge may have an existing schema for the conceptual information into which they can integrate the information from the computer lesson, experiencing less cognitive load because they do not have to process as much information in working memory when it is already incorporating in a long term memory schema. Alternatively, learners with more optics knowledge could experience higher cognitive load if they experience associated with the expertise reversal effect, which can occur when novices benefit from additional instruction but those with prior expertise suffer from lessons with too much detail (Kalyuga, Ayres, Chandler, & Sweller, 2003). These individual differences are controlled for in the current study by assessing the level of spatial ability and prior optics knowledge of the learners prior to the experiment (for more details, see Experiment in Chapter 6).

Literature

The theoretical frameworks of Embodied Cognition and CLT can be used to explain the cognitive mechanisms for the findings in previous literature that were used to develop the research questions investigated in the current set of studies. To understand the role of gestures in educational computer games, several disparate areas of research are synthesized below. First, Natural User Interfaces are explained to address the past and current state of gesture-based interactions in computer systems and how these interactions affect the user. One determinant of how the computer interaction affects the user is the mapping of the human-computer interface, or the relationship between an interaction and the object being controlled. Research on interface mapping is then discussed in the context of how closely the gesture-based interaction is to the learning material, or gestural congruency. This leads to the next area of research presented that highlights how gestures that are congruent with learning material may help learning conceptual information. By describing these areas of research, we can begin to see how gesture-based interactions in a computer lesson may affect the learner depending on how natural, or gesturally congruent, the interactions are to the learning material. Alternatively, the "naturalness" of an interaction may be subjective to the individual, and the way in which gesture-based interactions are perceived by and affect the learner may depend on how the interaction is instructed; therefore, previous research on the instruction of computer interactions is described to elucidate how medium of instruction may play a role in learning from gesture-based computer interactions.

Natural User Interfaces

As technology develops, human-computer interfaces have changed to meet the needs of interacting with more complex systems in user-friendly ways. Computer interaction today commonly consists of graphical user interfaces (GUIs) in which icons on the screen visually represent computer actions that can be selected and controlled by mouse input. GUIs that utilize features such as windows, icons, menus, and pointers have been referred to as *WIMP*s (Shiratuddin & Wong, 2011). In the mid-1980s, WIMPs replaced the command line interfaces (CLIs) of early computer systems in which the user typed commands via keyboard input to complete computer functions (van Dam, 1997). WIMP GUIs rapidly replaced CLIs as the mainstream computer interface, because these GUIs were seen as more user-friendly when learning how to interact with a computer system. In 1997, Andries van Dam, a prominent computer scientist and pioneer of computer graphics, described the eras of computer interfaces as "long periods of stability interrupted by rapid change," but expressed surprise 20 years ago that WIMPs had dominated user interfaces for so many decades (p. 63). Van Dam argued that a "post-WIMP" era of user interface can overcome limitations in the WIMP model by incorporating additional sensory modalities, natural language, or more than one user in control. He succinctly summarized the main problem with WIMP interfaces, stating:

> "However user-friendly, an interface is still an intermediary between the user's intent and execution of that intent. As such, it should be considered at best a necessary evil because no matter how fluid, it still imposes a layer of cognitive processing between the user and the computer's execution of the user's intent. The ideal interface is no interface (p. 64)."

Now, technological advances and dropping cost of motion tracking systems have opened the door to using gesture-based input to interact directly with computers, a form of natural user interface (NUI). The differentiating feature of NUIs compared to both GUIs and CLIs is that natural interfaces use the body as an input device to interact directly with the computer system, allowing the user to rely on existing skills of physical interaction (Roupé, Bosch-Sijtsema, & Johansson, 2014). NUIs can include gestures, speech, or touch to interact directly with the computer system. The purpose of NUIs is to provide the user with an interface that is easy to learn by not requiring much cognitive effort from the user. The term "intuitive" is used by human-computer interaction (HCI) researchers and product designers to describe the ease with which a new system is learned, and an intuitive user experience is a main goal of technology designers (Ullrich & Diefenbach, 2010). The extent to which NUIs are easy to learn and the cognitive mechanisms behind learning natural interfaces are a topical concern as these technologies become more prevalent.

In the domain of HCI, interactivity has been defined as, “the extent to which users can participate in modifying the form and content of a mediated environment…determined by the technological structure of the medium” (p. 84-85, Steuer, 1992). Unpacking this statement, it means that how the user controls the computer system depends on the structure of the HCI. One type of technological structure that allows the user to control computer actions is the computer mouse. Another type of input structure is gesture-based commands that are recognized by motion trackers. Various types of interface structures require different forms of interaction from the user. These interface structures can be described by their relationship, or mapping, between the type of input and the action they represent.

Continuum of Interface Mapping

There is a wide range in structures of computer interactions, existing on a continuum of “naturalness.” The relationship between a control and the object being controlled can be described as the degree of “mapping” (Norman; 2002). Steuer (1992) theorized a spectrum of mapping, from arbitrary commands that are not related to the action performed to “completely natural” commands that physically mimic the represented action. He continued by reasoning that the mapping of controls may be directly related to a real-world action (e.g., turning the hand clockwise to rotate a digital image in the same direction), controls may be metaphorical (e.g., scrolling down on a mouse to move downward on the screen), or controls may be merely arbitrary (e.g., double-clicking a mouse to select). Schwartz and Plass (2014) extended the ideas outlined by Steuer (1992), combining this proposed spectrum of arbitrary to natural mappings with a philosophy on types of mental action representations described by Bruner (1966).

Schwartz and Plass (2014) labeled three levels of interactivity to reflect Bruner's three levels of representation: 1. *Enactive mapping* occurs when the interaction closely resembles the physical action it represents, 2. *Iconic mapping* is when the interaction has features similar to the real-world action represented, and 3. *Symbolic mapping* is when the interaction is arbitrary and does not relate to a physical action (p. 245). This continuum of natural mapping is represented in Figure 1.

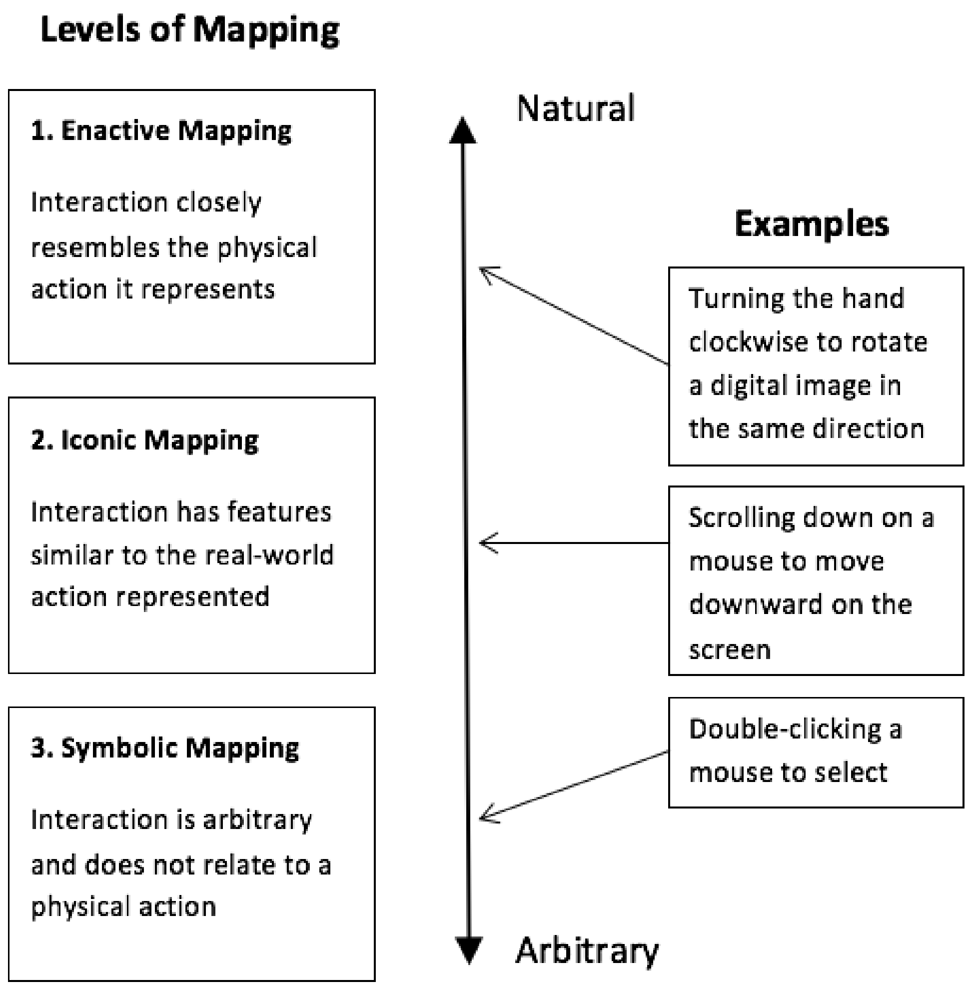


Figure 1. Schwartz and Plass' (2014) Levels of Mapping correspond to a continuum of mapping from arbitrary to natural, with Enactive Mapping being the most natural, followed by Iconic Mapping, with Symbolic Mapping being the most arbitrary. Examples of interaction for each type of mapping are described.

Schwartz and Plass (2014) contended that examples of these three levels of interactivity (i.e., Enactive, Iconic, and Symbolic mapping) are each common in human-computer systems, and they conducted an experiment to determine whether these descriptions are meaningful. The authors reasoned that these levels of mapping are meaningfully different if they differentially affect task performance or learning outcomes. The researchers compared how well participants could remember actions presented by either iconic or symbolic mappings of interaction, expecting the iconic condition to lead to better memory as it is a more natural level of mapping. In the iconic condition, participants completed a computer action by clicking and dragging to move an icon, representing the physical action of moving an object. The symbolic condition consisted of participants only clicking the icon to perform the moving task, demonstrating an arbitrary interaction. Participants were able to recall more actions and objects in the iconic condition (i.e., *drag*) than the symbolic conditions (i.e., *click*) in both immediate free recall (η_p^2=0.11) and delayed free recall three weeks later (also η_p^2=0.11); however, when the dependent variable was recognition instead of recall, participants did not recognize actions from the iconic condition significantly more than those from the symbolic condition at either immediate or delayed recognition. The authors concluded that actions in the iconic condition were recalled more than those in the symbolic condition. They conceded that the recognition results were inconsistent with this finding, but they justified this by stating that 1. Recall is harder than recognition and is therefore a more important learning objective, and 2. The conditions may have performed similarly due to potential issues with presentation of actions to be recognized.

Lending to the theory that more natural mappings are better for memory than less natural interactions, Schwartz and Plass' experiment (2014) showed that the distinction in levels of naturalness is meaningful because iconic computer interactions could lead to better recall than symbolic interactions. A major limitation of this study was that a fully naturally mapped, enactive condition in which the interaction closely mimicked the physical action represented (e.g., gesture) was not tested because only mouse-based input was utilized. The authors suggested that future research could incorporate motion tracking of body movements to create a more natural interface through enactive mapping. I am extending this work by Schwartz and Plass by directly comparing naturally-mapped and arbitrary gestures in a computer interaction task to assess how enactment affects learning a conceptual lesson.

A major issue of designing and evaluating natural user interfaces is defining what is "natural" or "intuitive." Although theories regarding natural mapping indicate that touchless, gesture-based technology can closely mimic real-world actions and lend to a feeling of intuitive interaction, until recently, appropriateness of gesture-based interactions was not a topic of much research (Grandhi, Joue, & Mittelberg, 2010; Sheu & Chen, 2014). The emerging research in this area is highlighting the issues researchers and designers face in understanding what it means for gestures to be "natural" and "intuitive." Gesture-based technology has been relatively limited in application to video games and research studies, and Grandhi et al. (2011) identified two challenges of designing gesture-based interactions that have limited their implementation: "1) achieving accurate and meaningful gesture recognition and 2) identifying natural, intuitive and meaningful gesture vocabularies appropriate for the tasks in question" (p.821). The first challenge relates to the limitations of technology in recognizing finer, more specific body movements that are more congruent to real-world actions. The second challenge arises from the issue of translating real-world actions to meaningful computer interactions, which is particularly challenging when what is meaningful may depend on personal experiences or cultural norms

(Abadi, Peng, & Zadeh, 2012; Mauney, Howarth, Wirtanen, & Capra, 2010). For example, Mauney et al. (2010) conducted a cross-cultural study of nine countries in which participants suggested touchscreen gestures to perform computer actions. Mauney and colleagues found that gestures were fairly similar across cultures when actions were less symbolic (i.e., pantomiming real-world actions), but Chinese participants preferred more symbolic gestures than participants from other countries. This is just one case in which considerations should be made for the intuitiveness of computer interactions based on user characteristics.

Due to these challenges in defining and creating natural gestures, most previous implementations of gesture-based commands were based on what was easier for motion-capture recognition (Nielsen, Störring, Moeslund, & Granum, 2004; Shiratuddin & Wong, 2011) or what felt less awkward for participants (Roupé, Bosch-Sijtsema, & Johansson, 2014); however, these *ad hoc* designs of gestures could cause unintended cognitive load, or mental effort, for the user and, "defeat the purpose of using gestures as a way to facilitate intuitive and natural interaction" (p.821, Grandhi et al., 2011). To my knowledge, the assertion that inappropriate (i.e., arbitrary) gesture-based interactions could increase cognitive load or mental effort during a task has not been empirically tested, and my proposed study can add to the literature of how natural gesture interactions compare to arbitrary gestures. The findings will lend evidence to which gesture-based interactions should be implemented in a conceptual computer lesson such that the gestures do not impose unintended mental effort on the learner.

Recent research has attempted to identify characteristics of gesture-based computer interactions that are natural and intuitive. Ullrich and Diefenbach (2010) explain that the term "intuitive" in the context of HCI is usually defined as the unconscious application of existing knowledge when using a new system; however, they argue that "intuitive" is a construct better described as four subcomponents that create an intuitive interaction: 1. Effortlessness, 2. Gut Feeling, 3. Verbalizability, and 4. Magical Experience. Unfortunately, these components are very loosely defined, to the point of being meaningless. Although Ullrich and Diefenbach introduce the idea of "intuitiveness" as a complex, multidimensional construct based on their understanding of the interaction literature, their exploratory study did not examine these components in any sort of factor analysis. More research would be needed to confirm these characteristics as unique factors, yet it may still be useful to consider that "intuitiveness" could be more than a single construct as presented in previous literature and should be clearly operationalized.

Furthermore, Ullrich and Diefenbach (2010) use the term "intuitive" to describe interactions that "make sense" to the user, but this term is more subjective than describing the interaction in terms of natural mapping to the real world or gestural congruency to learning material. As described later in this chapter, the subjective feeling of "intuitive" for an interaction may depend on how how the interaction is instructed, and not necessarily on the degree of mapping; therefore, the terms "intuitive" and "natural" may not be interchangeable. A gesture-based interaction may be naturally mapped to the learning material and also feel intuitive, or a gesture that is arbitrarily mapped may feel intuitive once it is instructed. For example, double-clicking a mouse is an arbitrary computer interaction because it does not match a real-world action nor does it necessarily relate to learning material if used in a computer lesson, but double-clicking a mouse might feel intuitive once it is instructed. The distinction between "natural" and "intuitive" may be further confused because researchers have investigated natural gestures in the context of a "natural feeling" that is subjective to the user and is not related directly to gestural congruency. Conceptualizing "natural" as a subjective feeling may be more related to the subjective feeling of “intuitiveness” than when “natural” refers to gestural congruency. To better understand whether participants perceive naturally mapped gestures as subjectively feeling “natural,” researchers have investigated what features of gesture-based interactions led to greater feelings of “naturalness.” Understanding whether naturally mapped gestures are perceived as natural by the user is also investigated in the current study to confirm that interactions with gestural congruency are also interpreted as feeling “natural” by participants.

Grandhi et al. (2011) conducted an experiment to determine what are characteristics of interactions perceived as “natural” by asking participants to enact scenarios that covertly represented computer functions. During the within-subjects experiment, participants were asked to gesture while explaining how they would complete a mundane action on an object, such as “close book” or “sort coins.” All actions involved acting upon or manipulating an object and some actions would require the use of a tool (e.g., “cut paper” would require scissors in real life). These mundane actions were chosen to represent computer functions without activating the participants’ previous experience with performing the computer tasks that may influence how they would gesture (e.g., “close book” represented the task of closing a computer function, and “sort coins” represented the computer task of arranging items). It was assumed that the more frequently participants used a particular gesture to describe these everyday tasks, the more subjectively natural that gesture must be.

Grandhi et al. (2011) were interested in what characteristics of gesturing were most likely to be used, and thus were more natural. The researchers were particularly interested in whether pantomiming the action was more natural than using the body to represent an object or tool. They reasoned that pantomiming is the more natural gestural interaction because developmental research indicates the typical way in which people represent actions in gestures is by acting as if holding an imagined object and not substituting the body for an object (Boyatzis & Watson, 1993). For example, when acting out cutting an apple, people are more likely to mimic holding the knife instead of using their finger to represent the knife. This is in line with the descriptions of natural mapping that suggest the closer the interaction is to the physical action it represents (i.e., pantomiming), the more natural and intuitive the gesture is. Additionally, the experimenters were interested in how frame of reference contributed to the naturalness of gesture-based interaction, hypothesizing that acting as if one is completing the action (egocentric reference) is easier than showing others how to do that action (allocentric reference).

To test these hypotheses about whether people find pantomiming more natural than using the body as a representation, and whether an egocentric reference is more natural than allocentric, Grandhi et al. (2011) varied: 1. Whether the participant was to respond either using an egocentric reference ("This is how I…") or allocentric reference ("You need to…"), and 2. Whether or not the participant was required to use his hand to represent a tool during the action. The results of the experiment supported the hypotheses, indicating that participants reported that it was easier to gesture an action by pantomiming the action from their own perspective, or egocentric reference. Participants were significantly more likely to pantomime an action than represent their body as a tool, and found it difficult to use their hand to represent a tool even in conditions when explicitly told to do so (participants were unable to do this 77.5% of the time). The experimenters concluded that for gesture-based interfaces to be natural and intuitive, they should be designed with an embodied approach in which the gestures are situated in physical experiences. In an attempt to prevent future *ad hoc* designs of gesture-based interfaces, Grandhi et al. (2011) provided a list of guidelines for developing natural and intuitive interactions to which I added examples for each guideline (Table 1).

Table 1. Design guidelines for natural gesture-based interactions (adapted from Grandhi et al., 2011)

Guideline	Example
1. Gestures should be dynamic representations of physical motion.	To rotate an object clockwise, the user should make a clockwise waving gesture.
2. Gestures should pantomime the action on an object or tool.	To select and move an object, the user should make a grasping gesture and then motion toward the target placement point.
3. Gestures should use both hands with the non-dominant hand situating the action of the dominant hand in space.	To perform an action on an object, the user should use one hand to hold the object and pantomime using a tool with the other hand.
4. Gestures should be conducted from an egocentric perspective.	To perform any action, the gesture should be from the perspective of the user acting on the object as if it were in front of him or her, and not from a viewer's perspective.

Although Grandhi et al. (2011) provide a much-needed set of guidelines for designing and researching natural gesture-based interactions, there are several limitations of this research. First, these guidelines were developed based on observations of gestures used to represent real-world actions. The authors did not create and test gestural interactions based on these observations, and testing could indicate usability issues, excessive mental effort, ergonomic problems, or technical limitations imposed by gestures that follow these guidelines. Additionally, the authors note that creating a generic set of guidelines that are appropriate for a broad number of tasks is inherently limited. These guidelines should be taken as a first step in creating NUI standards and should not be seen as concrete rules.

Several conclusions can be drawn from this summary of the natural user interface literature. Natural mapping is the extent to which computer interactions match real-world actions. The assumption is that the more natural and less arbitrary the mapping, the easier it is for users to interact with a computer system; therefore, interactions should be more natural. However, it is difficult to define natural in the context of computer interactions. The literature suggests enacting, or physically performing an action, is more natural than interactions that are more removed from the real-world or completely arbitrary. Gesturing is thus a logical way to create a natural mapping of the real-world to the computer interaction, although gestures can be more natural (e.g., pantomiming the real-world action) or more arbitrary (e.g., gesturing that is not related to a physical action). Because gestures can be either the highest level of natural mapping or the lowest, understanding how gestures help or hinder learning can guide how gestures should be incorporated appropriately in naturally mapped interactions.

Research on gesture-based interaction for training and education is relatively new, with few studies published before 2010 (for a meta-analysis, see Sheu & Chen, 2014). It is important to specify what is meant by “gesture” in this literature to systematically investigate what components of gesture-based interactions are best for various purposes. For example, Johnson-Glenberg, Birchfield, Tolentino, and Koziupa (2014) explained how touchscreen interactions have been referred to as “gestures,” and although previous research that found more naturally mapped touchscreen actions were better for learning a task than arbitrary touchscreen gestures, Johnson-Glenberg et al. question the use of the terms “gesture” and “embodied” to refer to touchscreen actions. They define a system as “embodied,” if the gesture-based interactions “activate multiple afferent and efferent neuronal pathways in the learner’s motor system” (p. 91). They argued that if the touch interaction does not reflect the content to be learned by mimicking the action it would have low “gestural congruency” and would therefore not be the best learning environment.

Johnson-Glenberg et al. (2014) stressed that research on gesture-based interactions should be more codified so that the term "embodied" in the context of learning is not "overused to the point of meaninglessness" (p. 89). To clarify what components of embodied learning environments (e.g., gesturing) are meaningful to learning outcomes and to provide a framework for studying embodiment in learning, Johnson-Glenberg et al. proposed the Taxonomy for Embodied Learning. See Figure 2 for a visual conceptualization of the taxonomy based on my interpretation of Johnson-Glenberg et al. This taxonomy is comprised of three continuous axes that represent characteristics of embodied learning environments on a continuum: 1. Motoric engagement, 2. Gestural congruency, and 3. Immersion. *Motoric engagement* involves how much the learner is able to move in a learning environment. Full-body movements and ambulation are at the higher end of the spectrum of motoric engagement because they entail more body movement, and clicking a mouse may be at the lower end of motoric engagement as it requires less body movement. *Gestural congruency* refers to how much a gesture in a learning environment corresponds to the content to be learned. This is related to the spectrum of natural mapping discussed earlier in the context of natural user interfaces, such that in an educational or training environment, enacted gestures that more closely represent the content to be learned are considered better for learning. Low gestural congruency could be arbitrary actions that do not relate to the learning material. *Immersion* refers to the perception of "being there" and a greater feeling of immersion is considered better for learning environments. Johnson-Glenberg et al. theorize that immersion is dependent on the technology used in the learning environment, and they suggest that head mounted displays may provide greater immersion than a small computer monitor and that does not occlude one's environment with a virtual world.

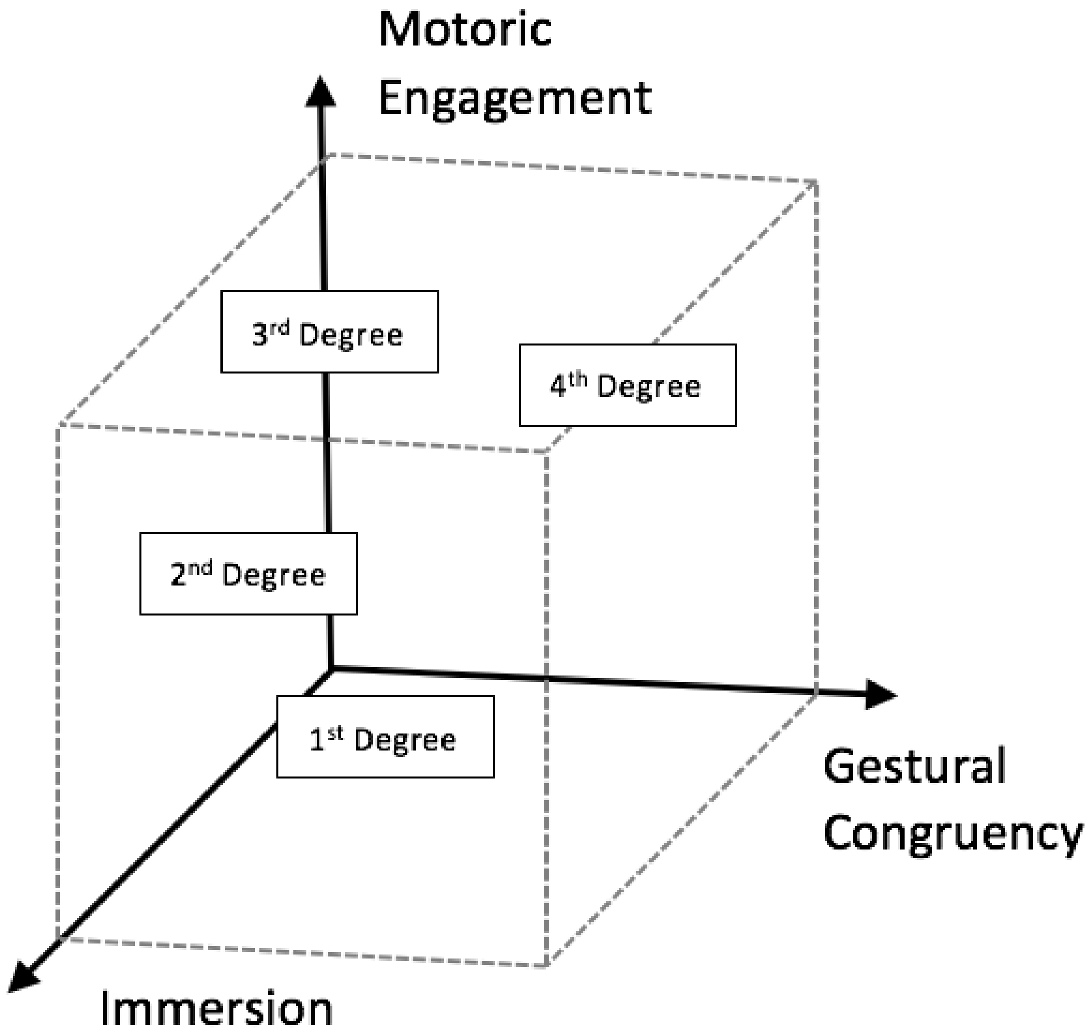


Figure 2. A visual interpretation of the Taxonomy for Embodied Learning based on the theory proposed by Johnson-Glenberg et al. (2014). The taxonomy consists of three continuous axes of embodied learning environments (1. Motoric engagement, 2. Gestural congruency, and 3. Immersion), in which higher levels of each component aid in embodied learning. The degrees correspond to how high on each axis an embodied learning environment may be, with the 4th degree representing the most embodied design and the 1st degree the lowest (see Table 2).

Although the three axes are conceptualized as continuous, the authors discretized these factors into four degrees (categories) to make more meaningful recommendations about embodied learning environments (Table 2). It should be noted that the edges of these degrees are not completely distinct as the degrees are meant to be a starting point to compare embodied technologies using a common taxonomy. Based on the descriptions of each degree in the taxonomy, the task used in the current experiment can be considered in the 3rd degree of embodied technology because it involves motion-capture of gesture-based interactions on a large screen monitor (semi-immersive).

Table 2. Four degrees of embodied technology (adapted from Johnson-Glenberg et al., 2014)

Degree	Technology	Motoric Engagement	Gestural Congruency	Immersion
4th Degree (highest)	Mixed-Reality; Ambulatory Motion-capture (e.g., SMALLab, Star Trek Holodeck)	Whole-body locomotion	Highly congruent gesturing with tangible or haptic manipulation	Highly immersive to semi-immersive
3rd Degree	Motion-capture and/or Large Display (e.g., Microsoft Kinect, Oculus Rift, HTC Vive, Flight Simulators)	Could be whole-body movement, but usually in one place	Highly congruent gesturing, but without tangible manipulation	Immersive to semi-immersive
2nd Degree	Interactive Small Screen (e.g., Desktop simulations and trainers)	Stationary	Congruent gesturing with interactivity	Not immersive
1st Degree (lowest)	Observational Small Screen (e.g., Educational videos such as Khan Academy, Crash Course, etc.)	Stationary	No gestural congruency nor interactivity	Not immersive

The current experiment focuses on the gestural congruency axis of the Taxonomy for Embodied Learning as it seeks to compare high and low levels of gesture-based mapping. The motoric engagement and immersion axes are not manipulated in the proposed study, but these are important factors to consider in future work on gesture-based interactions in various educational contexts. Johnson-Glenberg et al. (2014) emphasize that the boundaries for perceiving immersion are especially undefined in the potential interaction with embodied learning, yet these issues are beyond the scope of the current research questions.

Gesturing Helps Learning

Research on natural mapping for user interfaces points to gestural congruency with the real world as a way of increasing ease of use, but how does gesturing impact learning? Research conducted in the gesture literature can inform how gesture-based interactions will affect learning in a computer environment. By considering the ways in which gestures help enable processing of information into schemas, gestures that help learning can be implemented, while gestures that merely increase extraneous cognitive load can be avoided. The majority of the literature on the use of gestures in education focuses on how children learn by viewing teachers or parents gesturing to explain a concept or procedure (Singer & Goldin-Meadow, 2005; Vallotton, Fusaro, Hayden, Decker, & Gutowski, 2015), or by having the children produce either spontaneous or scripted gestures during a learning task (Cook, Mitchell, & Goldin-Meadow, 2008; Goldin-Meadow, Nusbaum, Kelly, & Wagner, 2001). The term "gesture" in these studies can refer to spontaneous gestures when communicating ideas (for a list of most commonly produced gestures parents make to scaffold children's learning, see Vallotton et al., 2015), or planned gestures made by teachers to clarify learning content (Singer & Goldin-Meadow, 2005; Vallotton et al., 2015). These studies have shown that both producing gestures and viewing gestured explanations during instruction results in better retention of knowledge.

It is important to note that the information taught in previous studies using gestures involves a wide-range of domains and types of knowledge, such as conceptual information (e.g., math concepts, Singer & Goldin-Meadow, 2005), problem-solving (e.g., Tower of Hanoi, block puzzles; Garber & Goldin-Meadow, 2002; Vallotton et al., 2015), and language development

(Capone & McGregor, 2004). The current set of studies investigates the role of gesture-based interactions for learning conceptual information (i.e., fundamentals of optics) from a computer lesson, so the following review of previous research focuses on learning conceptual information with gesturing, and how gestures may be particularly beneficial for conceptual knowledge. Specifically, gestures that are related to the learning content may benefit the learner by either lightening the cognitive load of the learner or activating the schema for a concept in the sensorimotor system. These theories are explored below with evidence from previous literature.

For example, Cook, Mitchell, and Goldin-Meadow (2008) investigated the role of gesture observation and production on learning conceptual information. In the experiment, the researchers asked children either to gesture or not gesture while explaining a new math concept. The children were presented with one of three ways of solving a math problem from the teacher: 1. Speech only condition in which the teacher did not gesture while explaining the concept, 2. Gesture only condition in which the teacher only gestured to explain the concept, or 3. Speech and Gesture condition in which the teacher both spoke and gestured to explain the concept. The teacher asked the children to repeat after her the instructions and/or gestures that she modeled during the explanation of the math concept. After this initial explanation, all children were presented with the same instruction incorporating both speech and gesture so that each student had the same overall information. The students then took a math test on the conceptual information immediately and then four weeks after the instruction. On the immediate post-test, all of the students improved on the math test without any significant differences between the three groups; however, performance on the delayed test four weeks later was better when the initial instructions included gestures (Gesture Only condition or Gesture with Speech condition) than when the initial instructions were only spoken (Speech Only condition). It should be highlighted that these results were based on task-relevant gestures, so the gestures and the content had high gestural congruency because the gestures and task were closely related. The authors concluded that gesturing played a causal role in learning the conceptual math knowledge, proposing the mechanism behind this is that gesturing is less mentally demanding when expressing information than speech alone, or gestures may facilitate better encoding by using enactment. These explanations are in line with the predictions made by Embodied Cognition and CLT theories in that the gestures alleviate cognitive load by off-loading processing from one modality (i.e., verbal) to another (i.e., motor), and the gestures were relevant to the learning material, which could facilitate germane processing related to creating schemas for that information as they direct attention to the key points.

The finding that gestures help learning leads to a question that is applicable to the natural mapping of computer interfaces: how similar must gestures be to the learning content to be beneficial? Although the previously discussed study by Cook et al. (2008) would suggest that gestures are useful when they relate to the educational content, some research indicates that gesturing during instruction helps most when the problem-solving strategies taught using gestures do not match the strategies explained in speech. Singer and Goldin-Meadow (2005) found that when a teacher gave gesture and speech instructions that included two different problem-solving strategies (mismatched gestures and speech) students learned more than when instructions were presented in either speech alone or speech with gestures. The authors believe this is because mismatched gestures with speech gave additional information to the student than a single strategy presented by speech alone or speech with matching gestures. In this context in which the mismatched gestures and speech give the student additional useful information for solving a problem, it seems feasible that incongruent gestures with speech aid learning because they were relevant for schema development, or germane load. The mismatched gestures in this case should not be confused with arbitrary gestures or gestures that are incongruent with learning content, which could have the opposite effect of increasing extraneous load. In the case of Singer and Goldin-Meadow (2005), the gestures did not match the accompanying instructions, but the gestures did relate strongly to the learning content. Because the mismatched gestures were highly relevant to the learning content, the mismatched gestures still had high gestural congruency in regard to the material to be learned.

Based on the framework of CLT, there are several possibilities for how gestures could affect learning: 1. Gestures could benefit learning if they reduce extraneous load and facilitate germane load – that is, gestures may lighten the cognitive load of the learner during a task by offloading the mental processing of the lesson from one modality (e.g., visual) to another (e.g., motor) – or 2. gestures could be detrimental if they increase extraneous load, such as arbitrary gestures that must be held in working memory at the same time the learning material is processed in working memory; therefore, determining whether gesturing is instructionally advantageous is a central concern. Previous research has investigated whether spontaneous gesturing increased or decreased cognitive load (Golding-Meadow et al., 2001). Goldin-Meadow et al. (2001)[1] reasoned that if gestures increase cognitive load during a cognitive task, memory would be worse than when not gesturing. Conversely, if gestures decrease cognitive load, memory during the cognitive task would be better when gesturing than not gesturing because gestures free up more working memory for other mental processing by offloading that mental processing onto another modality. The researchers conducted an experiment to test these hypotheses by giving participants (both children and adults) a list of items to be recalled followed by a math equation.

[1] Cognitive load is never specifically defined in Goldin-Meadow et al. (2001), so this explanation of cognitive load is inferred from the article and interpreted as Cognitive Load Theory (Sweller, van Merriënboer, & Paas, 1998).

Participants were instructed to explain how they would solve the math equation in either a gesture-allowed condition or a gesture-not-allowed condition. Afterward, participants were asked to recall the list of items presented before the math problem. Goldin-Meadow and colleagues found that more items were remembered from the list when participants gestured during the math problem. This finding was consistent for both children and adults and regardless of preexisting math knowledge, with the same proportion of items recalled for those who correctly or incorrectly solved the problem if they were allowed to gesture. The authors concluded that allowing for spontaneous gestures during a cognitive task can reduce the cognitive load imposed by the task. They suggest this is because gestures and speech are integrated in the limited resource system such that additional cognitive load on one modality (verbal or physical) can offset the effort needed to process the other modality. This explanation is in line with CLT and DCT in that mental processing can be offloaded from one modality to another, reducing extraneous load. Alternatively, the authors proposed that producing gestures might help organize information while speaking, thereby helping to conceptualize the information, which according to CLT would facilitate germane processing.

Another reason why gestures may help learning is that gestures may serve as activation for mental imagery, as predicted by the simulation theory under the Embodied Cognition paradigm. Hostetter and Alibali (2008) described gestures and their relationship to mental imagery in the *Gestures as Simulated Action* (GSA) framework, which explains the effect of gesturing based on simulation theory. Posited in GSA is that, “gestures emerge from the perceptual and motor simulations that underlie embodied language and mental imagery” (p. 502). The authors explained that spontaneous gestures are usually produced when people are describing their mental imagery or to express spatial and motor information. They argue that gestures are not merely epiphenomenal manifestations of the mental imagery being described, but instead gestures help facilitate spatial speech by activating the underlying motor representation of the mental image. Correspondingly, the imagined action simulates the physical action and a gesture results from the spreading patterns of neural activation, as suggested in simulation theory. Although the GSA framework describes gestures as they are produced spontaneously, it could be that this production of gestures with the activation of motor representations during the encoding or retrieval of mental imagery acts as a "cross modal prime." Gesturing may create stronger memories via embodied encoding and facilitate easier retrieval because the representation was coded in multiple modalities. Cross-modal priming could explain why spontaneous gestures during cognitive tasks help mental processing (as in Goldin-Meadow et al., 2001) or why gestures learned during a cognitive task led to better performance (such as Cook et al., 2008). This explanation is also in line with Hegarty, Mayer, Kriz, and Keehner (2005), who reasoned that the mental representations when solving a mental animation problem are inherently spatial, and spatial representations are easier to express physically (i.e., gestures) than verbally (i.e., speech). Hegarty et al. found mixed evidence that gesturing aided performance in a series of experiments in which participants were asked to determine the direction of a mechanical gear by imagining a gear sequence. The hypothesis that gestures are spatial as opposed to verbal was supported by their finding that people gestured to communicate information not included in speech. The hypothesis that spatial representations are more easily expressed in gestures also was supported by their results showing that people were more likely to gesture when also required to speak. Yet, although they found that people were more likely to gesture during speech and include additional information in gestures, Hegarty et al. found mixed evidence that gesturing aided performance on the mental animation problems, with no conclusive results that gesturing more helped problem solving. The reason for these lack of results seemed to be a split in the rate of spontaneous gestures produced by participants while they explained the mental animation problems such that some participants gestured on nearly every problem, but other participants gestured on almost no problems; however, the overall rate of gesturing was not correlated significantly with performance on the mechanical gear problems. They concluded that individual differences (e.g., spatial ability) in gesturing may account for the variability in performance and that gesturing may be more useful in mental processing for some people than others.

Although most research on watching gestures and producing gestures during problem solving or memory tasks has shown beneficial results, other studies have found that gestures may not always lead to the best problem solving strategy. Alibali, Spencer, Knox, and Kita (2011) conducted two experiments to test the effect of gesturing on strategy use in a problem solving task. The problem solving task involved predicting the movement in a series of gears. In the first experiment, participants were assigned to one of two conditions: 1. Gesture-allowed condition (with feet restrained) or 2. Gesture-not-allowed condition (with hands restrained). Participants were asked to explain how they would solve the gear movement problems, and responses were coded based on the strategies they used to explain their problem solving. Participants reported solving the problem using either a perceptual-motor strategy (e.g., mentally and/or physically simulating the gear movement) or an abstract strategy (e.g., knowing that an odd number of gears in a series will result in the last gear turning clockwise). The first experiment concluded that most participants in the gesture-not-allowed condition solved the task using an abstract strategy, while the gesture-allowed condition tended to use a perceptual-motor strategy (i.e., simulation). In a second experiment that also split participants into either a gesture-allowed or gesture-not-allowed condition[2], researchers found that the gesture-not-allowed condition was better at solving the problem than the gesture-allowed condition. For this particular task, an abstract strategy was more accurate than the perceptual-motor strategy. The authors concluded that gesturing influences strategy choice for completing tasks (such as facilitating a perceptual-motor strategy in the gear task), but the usefulness of that gesture-facilitated strategy choice depends on the task and may be detrimental or beneficial (i.e., in this task, an abstract strategy was more helpful than a gesture-facilitated perceptual-motor strategy). Although the majority of gesture literature seems to point overwhelmingly to beneficial results of gesturing on mental processing, gesturing was not the best strategy for problem solving in every situation.

Taken together, these experiments on gesture-based learning can explain why natural mapping and gestural congruency may be important for gesture-based interactions in HCI, but there are limitations on the extent to which we can extrapolate meaning from these studies on gesturing for learning. For example, producing spontaneous gestures is not the same as gesturing prescribed actions, such as gesture-based computer interactions, so the findings that spontaneous gestures reduce cognitive load and activate the motoric mental representations may not be true for non-spontaneous gesturing. Finally, many of these experiments did not address the naturalness or gestural congruency of the gestures with the learning content, but the degree of arbitrary or natural mapping of gestures to the learning material may be a key determinant of these results.

---

[2] No body parts were restrained in this condition of the second experiment, to reduce potential confounding effects.

Enactment Effect

Natural mapping of gesture-based interactions in a computer lesson may be more beneficial than arbitrary gesturing for learning new information because the learner is enacting the learning material when performing the naturally-mapped gesture, and physically performing information has been found to help recall for that learning material later (Nilsson, Cohen, & Nyberg, 1989). By the late 1980s, researchers investigating the role of different encoding modalities on memory were confident that physically encoding information about actions (such as the optics lesson used in the current experiment) is better than encoding in other modalities, stating "there is almost total consensus that enactment leads to higher recall levels through superior encoding" (p. 188; Nilsson et al.,1989). The finding that people retain more information when they physically perform the learning material is referred to as the *enactment effect*. Engelkamp and Jahn (p. 148; 2003) describe the enactment effect as a result of a multi-system (i.e., conceptual and sensory) account such that "conceptual information is enriched by sensory and motor information during encoding and retrieval." They suggest that the reason modality-specific encoding and retrieval is better depending on the type of the information is that the information medium directs which input (sensory) systems are activated and the information is not stored just conceptually. The authors describe how an action phrase may activate both the conceptual system (verbal information) and the motor system (physical information), thereby encoding the action phrase in multiple ways. Engelkamp and Jahn further suggest that the sensory and/or motor information that was encoded can be "regenerated" when the information is retrieved, which fits with the simulation theory, leading to better recall when the encoded material is reactivated because the memory traces are stronger. The authors claim the multi-system account is supported by evidence showing that more complex stimuli containing more detailed sensory information is remembered better as well as evidence from brain imaging studies showing reactivation of sensory areas during recall of enacted information.

To support this theory of a multi-system enactment effect, Engelkamp and Jahn (2003) conducted two experiments on memory of verbs and objects after enacted (physical) or verbal encoding. In the first experiment, German university students memorized sets of verbs and objects either by reading the lists (verbal) or by reading and enacting the verbs (verbal and physical). The lists of words were also manipulated by the strength of association between the verbs and objects (either weak or strong associations between the verb and object), and were manipulated by the structure in which the words were presented (either lists or phrases containing the verbs and objects). The strength of verb/object association and the presentation structure were included as variables because an interaction between either of these factors with enacted encoding would indicate that the enactment effect is only as strong as the conceptual structure underlying the encoded information. For example, if enacted encoding was only better than verbal encoding when the verb/objects were strongly related to each other, this would indicate the conceptual system is needed in conjunction with physical encoding. Participants were asked to recall the verb/object combinations they had memorized, and the results indicated that more was recalled after enacted encoding than verbal encoding, supporting the enactment effect. Additionally, more verb/object combinations were recalled for strongly associated verb/objects than weakly associated, and verb/object combinations presented as phrases were remembered better than those learned as lists, but there were no significant interactions of these three factors. Because there were no significant interactions between the type of encoding and other factors (specifically, enacted encoding was better regardless of strength of verb/object association or phrase/list presentation), Engelkamp and Jahn concluded that these findings support a multi-system enactment effect, because: 1. Enacting the verbs (i.e., motor system) led to better recall and was not dependent solely on the strength of relationship between the verb/object (i.e., the conceptual system), and 2. The structure of the information (list or phrase) was also independent of encoding modality, indicating the effect of engaging the motor system is useful beyond the conceptual structure of the information. The authors' second experiment replicated the first, but cued recall was used to assess memory instead of free recall to test directly list or phrase presentation on enacted recall. The results mirrored those of the first experiment, supporting the enactment effect by showing that enacted encoding was better for cued recall. Strongly associated word pairs were again better recalled than weakly associated pairs, and this again was independent of encoding modality. Recall for the list or phrase presentation, however, was not significantly different when the recall was cued. This result was expected because the structure of the word pairs would not be as important for recall when a cue was given. Overall, these experiments taken together support the enactment effect for remembering verbs and objects that is independent of how the words are related or structured.

Furthermore, the enactment effect may produce stronger memory traces over time than verbal instruction. In fact, performing an action once is better for memory than seeing and hearing verbal instructions twice, even after a week. Nilsson, Cohen, and Nyberg (1989) conducted three experiments to test how encoding action phrases in different modalities affected forgetting over time. They compared subject-enacted (physical) encoding, in which the participants performed action phrases they were asked to remember, to visual and verbal encoding where the participants saw the action phrases on cards along with the phrases spoken by the researcher (visual and auditory encoding of verbal information). Participants were asked to recall the action phrases they had encoded either physically or verbally both immediately after presentation and at delayed intervals ranging from a few minutes to seven days. The three experiments varied the recall intervals and the number of presentations for each type of encoding modality, and found converging results. All of the experiments found that enacted encoding by physically performing the action phrase was better for memory at each recall stage, and this was true even when the verbal encoding was presented four times compared with only one presentation with enacted encoding. Additionally, Nilsson et al. found that the slopes of forgetting over time were the same for both enacted and verbal encoding such that, while both modalities showed declines in recall with each subsequent interval, memory for enacted phrases remained better than verbal phrases at the same rate. Because the rates of forgetting for both modalities were consistent, the authors postulated that a similar mechanism is responsible for physical and verbal encoding. This set of experiments has clear implications for the proposed study, suggesting that when instructions include physically performing an action to be learned (e.g., gesture-based computer interactions), enacting the learning material will lead to better memory than verbal encoding, and these learning outcomes should last longer over time. The current study extended this theory by testing whether the enactment effect was due to the physical enactment of the learning material (i.e., natural gestures), or whether any engagement of the motor system can create and enactment effect, such as arbitrary gestures. Finally, if natural gestures are beneficial, are they beneficial because they are tied to the learning material, or does the way in which the gestures are instructed impact how natural the gestures are perceived?

Instruction of NUIs

The ease of interacting with a human-computer system could be due to the "naturalness" of the mapping between the interaction and the real-world; yet, an alternative hypothesis may be true: The ease of interaction may depend on how well an interaction is instructed or trained. If gesture-based interactions are instructed such that they activate the motor system, the interactions may be easier to learn regardless of how well they match the learning material, resulting in feelings of naturalness (Borghi, 2007). For example, arbitrarily mapped computer commands that are common today, such as double-clicking a mouse to select an item on the computer screen, at some point had to be instructed. Once a computer interaction is learned, it may be easy to use or feel intuitive regardless of the naturalness of the interaction, so we should consider effective ways to instruct computer interactions to overcome limitations of arbitrary mapping. Because there is not yet a standard vocabulary for gesture-based interactions, and gesture designs can range on a continuum of mapping (including arbitrary gestures), appropriateness of different instructional methods for teaching gesture-based interactions may be important when the interactions are not inherently intuitive for all users. In this section, the most effective instructional strategies for learning novel computer interactions is discussed.

Medium of Instruction

Looking at the literature on instructing computer interfaces, the sense modality in which gesture-based interactions are instructed can differ in the type of media in which instructions are presented, such as video or text-based tutorials. The medium of instruction may influence how users perceive the interaction as better or worse for understanding the gestures and interacting with the computer lesson. Better instructions would be those that make interactions seem easier, thereby reducing extraneous cognitive load of the interaction. Worse instructions would be those that make the gestures not make sense when interacting with the computer lesson, increasing the extraneous processing of information not associated with the learning material as the user must process how to gesture instead of the conceptual lesson. For example, researchers have argued that instructing new user interfaces using video-based instructions is better than verbal instructions, because videos can help the user to visualize hard to imagine actions, such as human movement, and they encourage multi-modal processing (Alexander, 2013; van Gog, Paas, Marcus, Ayres, & Sweller, 2009). From a theoretical perspective, understanding the gesture-based interactions may be affected by the modality in which the interactions are instructed. The Embodied Cognition approach would suggest that instructions that engage the motor system may facilitate better encoding and retrieval of the gestures than instructions that do not prime the motor system. In the current experiment, this would mean that video-based tutorials that depict the gestures being performed could activate the mirror neuron system for those actions. With the sensorimotor system primed for those gesture actions, mental representations for the gestures can be stored in the sensorimotor system and can later be recalled in the same sensorimotor state. Text-based tutorials on the other hand, may not activate the sensorimotor system, and may instead be processed as verbal information without the benefit of priming the sensorimotor system. Learners with text-based instructions may store the gesture instructions as verbal information and would therefore not have the advantage of visualizing the gestures, activating the neural system for those gestures, and storing the gestures as sensorimotor information. Then, when the gestures are used in the computer lesson, the instructions for the gestures may be recalled as verbal information, resulting in a mental simulation of verbal information instead of a visualization for the gesture. It may be harder to form mental representations and recall sensorimotor instructions for the gestures (i.e., visualizing the action) when the motor information (i.e., gesture action) is presented in an alternative modality (i.e., verbal).

Alternatively, the modality of instruction might impact the load associated with the gesture-based interactions. From a CLT perspective, Sweller, van Merriënboer, and Paas (1998) explained that modality effects derive from the cognitive load associated with integrating sensory information in working memory, such that the incoming sensory information can overwhelm the amount of processing with too much information in one modality. They describe that the underlying premise of major working memory models is that sensory information is processed in different subsystems (i.e., auditory, visual), and each subsystem can become burdened by too much sensory information to process[3]. Sweller and colleagues reasoned that this is due to a split attention effect, such as when several pieces of verbal information must be processed simultaneously (e.g., written text and spoken text), and the verbal processing subsystem must integrate different pieces of information in working memory, dividing the attention on each piece of information. The way to alleviate the cognitive load on each subsystem is to present information in different modalities so as not to overwhelm a single modal processing system. In the current study, presenting gesture-based interactions in text-based tutorials might have the benefit of offloading the processing onto different modalities. Gesture-based interactions can be considered sensorimotor information that may be mentally processed in a nonverbal cognitive structure, whereas a text-based tutorial of instructions would be verbal information. By processing the gesture interactions (i.e., nonverbal information) with text-based instructions (i.e., verbal information), it may be that neither system is overwhelmed with too much information to process at once. On the other hand, text-based tutorials may not be more beneficial than video-based tutorials if the video instructions reduce extraneous load. Video-based tutorials for interactions could reduce extraneous processing in learning the gesture-based interactions because they help the user visualize what the gesture looks like, highlighting the key features of

3 There is a more narrowly-defined modality principle under the Cognitive Theory of Multimedia Learning (CTML) as described in Mayer and Moreno (2010), which specifies that spoken narration, and not written text, should be used in conjunction with visual information in instructional systems so neither modality is overwhelmed. This specific form of modality effect is reviewed in Ginns (2005). The term *modality effect* is used in a broader sense in this paper to mean the effect of one modality as opposed to another in mental processing.

the gestures (e.g., shape of the movement, starting and stopping locations) that may be less salient in text-based instructions. Previous research on how modality of instruction affects learning can give insight into what kind of tutorial would be best to teach the gesture-based interactions for the current computer lesson.

Previous research has investigated methods of instructing interface interactions in different modalities (e.g., visual, verbal) and media (e.g., video, text). Schurmann, Binder, Janzarik, and Vogt (2015) tested whether the intuitiveness and usability of a multi-touch technology depended on the instructions for the interactions. In this study, multi-touch interactions were touch-based gestures on a touchpad, such as a pinching motion with the thumb and pointer finger to "shrink" an object on the screen. Participants were assigned to one of three instruction conditions to learn the touchpad gestures: text-based instructions, video-based instructions, or a control condition with no instruction. Participants then completed a task using the touchpad gestures and then rated their perceived usability of the gestures. Although all participants learned touchpad gestures that were intended to be intuitive, the researchers found that more touchpad gestures were used when they were instructed versus no instruction, and video-based instructions influenced the perception of the quality of the gestures. Schurmann et al. suggested these results are due to video instructions providing information that "bypass[es] conceptual difficulties" associated with learning new interaction methods (p. 251), and that the touchpad gestures themselves were not intuitive. The authors concluded by stating product developers should provide video-based tutorials for interacting with new systems because, "while near to every electronic product arrives with some kind of text instructions…these may not differ from providing no user instruction at all" (p. 255).

Another study investigated whether instruction of computer tasks influenced understanding of and memory for the computer tasks. In Alexander (2013), a word processing computer task was instructed using either video- or text-based instructions. The author predicted that the video instructions would provide better understanding of the computer task because the videos could illustrate hard to visualize information like human movement, as well as utilize multimodal processing. The results indicated that video-based instructions of the computer task were better in terms of higher accuracy during the task and higher recall of the task afterward than the text-based instructions. Users also rated the video instructions more favorably than text instructions on four Likert-scale measures of satisfaction: 1. Level of comfort, 2. Ease of use, 3. Ease of remembering, and 4. Overall usability. Because users made fewer errors on the computer task with video instructions, the author concluded that videos facilitated better understanding of the computer task. Alexander also summarized guidelines from the literature for developing video-based instructions for user interfaces. Video instructions, she stated, should walk the user through the interface by: 1. Chunking information into small pieces, 2. Highlighting important information, 3. Contain information about goals of the computer task, and 4. Be understandable to a variety of users.

Although some research has found that video-based instructions can help learning computer tasks, other studies have not found an advantage of videos over other media as an instructional technique. Although not directly testing instructional techniques for computer tasks, Mayer and Anderson (1991) tested the more general question of whether animations in a computer lesson increase learning more than verbal information. The researchers tested whether animations were better for a learning a complex conceptual task compared to the information presented verbally in narrated speech. In two experiments, they found that neither the animations alone nor the narration alone helped problem solving or recall; however, animations presented with narration were better for both problem solving and recall. This research comparing animations and verbal information in a computer lesson can be applied to the more specific question in the current set of studies investigating type of instruction for interacting with a computer lesson. Because Mayer and Anderson (1991) did not find animations to be beneficial as an instructional technique compared to verbal information in a computer lesson, the current study might also not find the predicted benefit of video over text instructions as video instructions are structurally similar to animations (i.e., dynamic visual information) and may have the same effect on learning and cognitive load in a computer lesson. In a review of similar studies, Ayres and Paas (2007) explained why the theorized benefit for animations and videos were not found in some studies. Although CLT theories, as previously discussed, predict videos reduce extraneous load and facilitate germane load by directing attention to relevant information, Ayres and Paas reasoned that video instructions could be a distraction from subtler features of the learning material, and in fact increase extraneous load. They explored another reason that videos might not show the hypothesized advantage over other media in that videos are transitory, and new information must be processed while previous information may not yet have been processed. To mitigate the issue of the transitory nature of videos and animations, the authors suggested adding user control such that the video can be paused or reviewed to process information at the learner's pace. Additionally, the ability to segment the video instructions in this way may facilitate germane processing, thereby achieving the intended benefit on cognitive load. The notion that self-pacing moderates the relationship between modality of instruction and learning has been supported in other research (for a review see Ginns, 2005), so the current experiment allowed for self-pacing of both video- and text-based instructions in that videos could be paused or rewatched and all slides of the tutorials could be reviewed. Based on the research that finds the method and modality of instruction may impact learning or perception of new computer interactions, the way in which the gesture-based interactions were taught was manipulated in the experiment.

CHAPTER THREE: THE CURRENT RESEARCH

To determine whether type of user interactions or the way in which the interactions are instructed affects learning conceptual information in a gesture-based computer lesson, the current research investigated the effects of naturally mapped or arbitrary gestures in conjunction with video- or text-based instructions. The first two studies established what specific gesture-based interactions participants produce and rate as most natural for the task, which determined what gestures are appropriate for the experimental testbed. In the main experiment, to parse out the effects of natural mapping, type of instruction, or their additive effects, a two-factor design was used to elucidate the relationship between these instructional techniques. The first factor was type of gesture-based interaction (naturally-mapped or arbitrary gestures) and the second factor was type of instruction (video or text).

The extent to which a computer interaction corresponds to the action it represents on the screen can range from arbitrary to natural mapping. Previous experiments that address natural mapping either ask participants subjective questions about the intuitiveness of the interaction (Schürmann et al., 2015; Silpasuwanchai & Ren, 2015), or they give categorical labels to the level of naturalness based on theoretical reasoning (Nielsen, Störring, Moeslund, & Granum, 2003). To quantify the subjective nature of feelings of naturalness, the first two studies were conducted to address the questions: What gestures do people think are natural when performing common object manipulation actions (Study 1), and are the produced gestures then interpreted as natural or arbitrary by other people (Study 2)? After developing the gesture-based interactions systematically and then piloting these gestures in the testbed, manipulation check questions following the main experiment asked participants how they would rate the gestures' degree of naturalness to ensure the manipulation was salient. These studies provide confidence that the gestures included in the natural and arbitrary gesture conditions are appropriate for their respective conditions.

This set of studies adds to the literatures on natural mapping and enacted instruction by parsing out the effects of these theories on performance and learning in a gesture-based computer lesson. In the main experiment, participants used either the naturally-mapped or arbitrary gesture-based interactions after receiving either video- or text-based instructions (2X2 between-subjects design). The computer lesson involved learning conceptual information about the fundamentals of optics by manipulating lenses and mirrors in a beam of light using gesture-based interactions with the computer lesson. Manipulating the lenses and mirrors in the beam of light resulted in refraction or reflection of the light, illustrating to the participant how different lenses and mirrors change a beam of light and result in an altered image of the object being reflected/refracted. After the computer lesson, the instructional techniques were assessed by comparing how much conceptual information was learned and the amount of mental effort expended from the lesson. In this 2X2 design, effects of interaction and instruction on conceptual knowledge can be studied. Results suggest the extent to which natural mapping and/or enacted instruction are effective for producing learning outcomes on a conceptual task.

Results of this set of studies will inform the research questions of whether more naturally-mapped gestural interactions are better for learning from a computer lesson than arbitrarily-mapped gestures, and whether type of instruction for the gesture-based interactions can influence the computer lesson.

Hypotheses

Based on the theoretical frameworks of Embodied Cognition and CLT, as well as the empirical evidence from previous research, the following hypotheses were predicted to answer the research questions.

Hypothesis 1

The first research question asked whether the type of gesture-based interaction affects learning and cognitive load in a conceptual computer lesson, and is addressed with Hypothesis 1. Because previous research has found that gestures that are naturally mapped to the learning material help a learner understand and remember that information, I predict that interacting with the computer lesson using natural gestures will lead to higher instructional efficiency for that instructional technique than using arbitrary gesture interactions.

H1: Natural mapping of gesture interactions will lead to better learning and lower perceived cognitive load, producing higher instructional efficiency for natural gesture interactions than arbitrary mapping of interactions.

Hypothesis 2

The second research question asked if the medium with which the gestures are instructed affects learning and cognitive load on a computer lesson. As predicted with Hypothesis 2, video-based tutorial instructions for the gestural interactions are expected to be a better instructional technique than text-based instructions. This is supported by the previous research that indicated video-based instructions may help the learner to visualize the gesture-based interactions by activating the sensorimotor system for that motor information.

H2: Video instructions of the gesture interactions will lead to better learning and lower perceived cognitive load, producing higher instructional efficiency for video-based instructions than text-based instructions.

Hypothesis 3

The final research question asked whether there would be combinatorial effects of type of gesture interaction and medium in which those gestures were instructed. Results from the experiment may show that there are only differences in instructional efficiency when type of instruction and interaction are combined. For example, using natural gestures to interact with the computer lesson may be more instructionally efficient than arbitrary gestures only when the interactions are instructed using a text-based tutorial, because it is harder to visualize gestures that do not make sense for the computer lesson from text rather than videos. Alternatively, natural gesture interactions that correspond with the learning material may be more instructionally efficient than arbitrary interactions regardless of how those interactions are instructed because the natural mapping reduces cognitive load and helps schema development. The prediction for this hypothesis is that there will be combined effects for type of gesture-based interaction and medium of instruction such that natural gestures with video-based tutorial will be more instructionally efficient than the other conditions, and arbitrary gestures with text-based tutorial will be less instructionally efficient than the other instructional techniques.

H3: Natural mapping with video instruction together will lead to the highest learning gain and lowest perceived cognitive load of all the combinations of mapping and instruction, resulting in highest instructional efficiency, while arbitrary mapping with text instruction will be the worst for learning with and highest perceived cognitive load, resulting in lowest instructional efficiency.

Potential Confounds

The main constructs of interest for this experiment are the extent to which using natural gestures to interact with a computer interface helps learning conceptual information and whether type of instruction affects this relationship. In the course of testing these two constructs in the experiment, potential confounds can be anticipated that would directly or indirectly affect performance on the task and subsequent measures. Several individual differences may play a role; in particular, a participant's preexisting knowledge of optics, his or her spatial ability, and whether he/she has experience with video games may contribute to differential effects or interact with the constructs of interest. Next, potential confounds are described, along with ways of mitigating or controlling for these effects.

Knowledge of Optics

Knowledge of optics and physics concepts, particularly how light waves interact with lenses or mirrors, may directly affect scores on the Knowledge of Optics post-test measure that is intended to evaluate knowledge gained from the learning material in the experiment. A participant with prior knowledge of optics would likely perform better on a test of said knowledge than one without familiarity of that domain. There may also be an interaction effect such that participants without much knowledge of optics may gain more from the learning material in the experiment than those who already have more knowledge of those topics. In terms of cognitive load, those with prior knowledge may report less mental effort because they have an existing schema for the conceptual material so they may not need to process as much new information in working memory. Alternatively, participants with more prior knowledge may feel more cognitive load and learn less from the lesson if they experience an expertise reversal effect, which occurs when those with higher expertise are hindered by too much information presented in a lesson while those with less knowledge benefit from more detailed information (Kalyuga et al., 2003). The difference between pre- and post-test scores on the measure of optics concepts may be greater for those who started with less knowledge.

The confound of prior knowledge of optics and physics concepts can be addressed by giving a pre-test measure of optics during the online prescreening portion of the study and then removing participants from analyses who score more than the typical range of participants (e.g., two standard deviations above the mean score). Additionally, administering the Knowledge of Optics measure during the online prescreening prior to the experiment will reduce the likelihood of priming optics and physics concepts immediately before the learning material is presented in the computer lesson.

Spatial Ability

Spatial ability may play a confounding role in this experimental task, which requires participants to develop an understanding of optics concepts while coordinating their body movements to the movement of objects on a computer screen (for detailed description of the computer lesson, see Experimental Testbed subsection below). The concepts presented in the computer lesson are highly spatial in that the participants must manipulate lenses and mirrors to different orientations to observe how the angle of the light waves change as a function of the lens/mirror interaction. The gesture-based interface is also a spatial task as it necessitates that the participants track their physical movements in space as they correspond to the actions displayed on screen. A participant's ability to visualize differing viewpoints may therefore be of interest.

Spatial ability will be measured because it may contribute to performance on either the post-test of physics concepts or performance on the computer task itself (i.e., gesturing). To clarify, higher spatial ability may help a participant to score better on the optics measure that includes questions on the effect a lens/mirror has on the angle of a beam of light (Hegarty & Sims, 1994). Interacting with the computer lesson via gestures is also inherently spatial, involving the mental mapping of bodily movement in space to corresponding changes in the lens/mirror on the screen; therefore, the spatial task of moving the on-screen objects may be performed better by those with higher spatial ability regardless of experimental condition. Additionally, there may be an interaction of spatial ability with one or both of the manipulated variables (i.e., type of gesture or type of instruction). A meta-analysis identified that higher spatial ability is related to better learning from dynamic visualizations (i.e., animations or videos), so participants with high spatial ability in the current study may perform better when given video-based tutorials than those with low spatial ability (Höffler 2010). Or, those with lower spatial ability may benefit disproportionally from either natural mapping or video instructions because those with higher spatial skills might already perform the task better and therefore have a lower gain score on outcome measure (i.e., optics post-test). The potential confound of spatial ability will be controlled by measuring spatial ability in a prescreening questionnaire using the Paper Folding Test (described in the Materials section below). The Paper Folding Test was chosen because it measures the spatial component of spatial visualization, which involves mental rotation while holding additional pieces of information in working memory, similar to the conceptual information of the optics lesson in which the learner must visualize how a rotated lens or mirror will reflect light according to the lens or mirror properties. Measuring spatial ability and using it as a covariate will allow for using statistical methods during data analysis to partial out the effect of spatial ability on conceptual learning.

Video Game Experience

Experience with video games may affect the learning of a novel computer-based task, either positively or negatively. Differential effects for those with previous experience with video games may occur such that skills in video games transfer to new computer environments. For example, people with previous video game experience may do either better or worse on the current computer task because their expectations do or do not match the experimental task, so it may take more or less mental effort to adjust expectations to meet the current task demands. To address the potential confound of video game experience, participants will be asked about prior video game experience in the online prescreening questionnaire prior to the in-lab study. Video game experience can then be accounted for in the variance of outcome measures (i.e., optics test, mental effort scale) during statistical analyses by including video game experience as a covariate.

CHAPTER FOUR: STUDY ONE

What are Natural Gestures?

One of the main research questions addressed by this set of studies is whether natural gestures are better than arbitrary gestures for learning a conceptual task when using a gesture-based computer interface. To test this question, the type of gesture (natural or arbitrary) participants used to interact with the computer lesson was manipulated in the main experiment. To determine what gestures were considered natural and arbitrary by participants, two studies were conducted to develop and confirm the gestures used in each condition. The first step to this was identify what natural gestures would be produced spontaneously by participants for the computer interactions. The goal of the first study was to down-select the potential gestures considered natural for nine common manipulations for computer interactions that were used in the experimental testbed (Table 3).

In the final experimental testbed, participants used the gestures determined in the first and second studies to manipulate lenses and mirrors in a computer-based physics lesson. To determine what specific gestures are considered natural by users, I first restricted the number of natural gestures for each action to the most likely candidates in Study 1. Participants were asked to perform a gesture to indicate a desired action to a new computer system. A series of common computer actions were presented to the participants, and their gestures were recorded using an infrared motion tracker. Results of Study 1 were gestures that are considered natural for computer interactions by the participants who produced them, and these gestures were subsequently rated by a separate group of participants in Study 2 on their perceived naturalness.

Table 3. List of Computer Actions for Experimental Testbed

1. Moving an object up
2. Moving an object down
3. Moving an object left
4. Moving an object right
5. Rotating an object clockwise
6. Rotating an object counterclockwise
7. Selecting an object
8. Enlarging an object
9. Shrinking an object

Method

Participants

Participants (*n*=17) were students recruited from the university research participation pool who received class credit for completing the study. Ten were female, and seven were male. The ages ranged from 18-20 years old (*M*=18.31, *SD*=0.60), and all were predominantly right-handed. Participants who were not predominately right-handed were excluded from participating in the current study because the main experiment was limited to right-handed individuals and gestures were therefore developed for those who were right-handed. The limitation of using right-handed participants was due to two factors: 1. Gestures that are natural for right-handed individuals may not be perceived as natural for left-handed individuals or performing right-handed gestures may induce more cognitive load for left-handed individuals, thereby confounding the manipulation of gesture mapping, and 2. The computer lesson testbed used in the experiment implemented a set of 18 gestures that were naturally and arbitrarily mapped to the nine actions, and including a second set of gestures for left-handed individuals would have doubled the amount of programming needed to develop the testbed. Therefore, due to concerns related to the experimental design and testbed development, only right-handed participants were included in the current study.

Procedure

Participants signed up to participate in the in-lab study through the university research participation system. The research participation system recorded in its prescreening the three demographic questions used: age, sex, and dominant handed-ness. Once in the lab, participants were given an informed consent form. After reading an informed consent and agreeing to participate, participants were asked to stand on a mark on the floor facing a Microsoft Kinect V1 infrared motion tracker. The Microsoft Kinect was used to record video and to capture depth information and joint coordinates as participants performed gestures. Participants were then asked by the experimenter to perform gestures to show how they would interact with a computer to complete a series of actions. The series of actions included common object manipulations to be used in the experimental testbed, and the nine actions were given to participants in a random order (Table 3). After performing a gesture for each action, participants were given a post participation information sheet to debrief the study. Participants took approximately fifteen minutes to complete the study. The videos of the gestures recorded by the motion tracker were then analyzed for converging features (e.g., starting height of the action, movement shape, and direction) to determine the most natural gestures for each action.

Results

The gestures for each action were classified based on characteristics of naturalness (Table 4). The most often gestured features of each characteristic are shown in bold. The coding scheme was adapted from Grandhi et al. (2011) in which features of gestures were analyzed to determine characteristics of "naturalness." The features coded in Grandhi et al.'s analysis included: 1. Whether gestures were right or left handed, 2. Whether one or both hands were used, 3. If the gesture was pantomimed or the body was used as a tool or object, 4. Whether the gesture

Table 4. Characteristics of Natural Gestures Produced for Each Action

	Gesture Characteristics				
Action	**Handedness**	**Number of Hands**	**Pantomime or Body-as-Object**	**Static or Dynamic**	**Main Action**
1. Up	**70.59% Right** 11.76% Left 17.65% Both	**82.35% One** 17.65% Both	**76.47% Pantomime** 23.53% Body-as-Object	100% Dynamic	100% Main
2. Down	**94.12% Right** 0% Left 5.88% Both	**94.12% One** 5.88% Both	**70.59% Pantomime** 29.41% Body-as-Object	100% Dynamic	100% Main
3. Left	**70.59% Right** 11.76% Left 17.65% Both	**82.35% One** 17.65% Both	**64.71% Pantomime** 35.29% Body-as-Object	100% Dynamic	100% Main
4. Right	**88.24% Right** 5.88% Left 5.88% Both	**94.12% One** 5.88% Both	**64.71% Pantomime** 35.29% Body-as-Object	100% Dynamic	100% Main
5. Clockwise	**82.35% Right** 5.88% Left 11.76% Both	**88.24% One** 11.76% Both	**64.71% Pantomime** 35.29% Body-as-Object	100% Dynamic	100% Main
6. Counter-clockwise	**82.35% Right** 5.88% Left 11.76% Both	**88.24% One** 11.76% Both	**64.71% Pantomime** 35.29% Body-as-Object	100% Dynamic	100% Main
7. Select	**94.12% Right** 5.88% Left 0% Both	**100% One** 0% Both	23.53% Pantomime **76.47% Body-as-Object**	100% Dynamic	100% Main
8. Enlarge	23.53% Right 0% Left **76.47% Both**	23.53% One **76.47% Both**	**82.35% Pantomime** 17.65% Body-as-Object	100% Dynamic	100% Main
9. Shrink	29.41% Right 0% Left **70.59% Both**	29.41% One **70.59% Both**	**88.24% Pantomime** 11.76% Body-as-Object	100% Dynamic	100% Main

Note: Coding was based on the characteristics of natural gestures defined by Grandhi et al. (2011). The most often gestured feature of each characteristic are shown in bold.

was static or dynamic, and 5. Whether the gesture referred to the main action (hand pantomiming cutting) or was used to set up the context of the gesture (hand pantomiming holding the object being cut)[4].

In the list by Grandhi et al. (2011), some features needed further clarification because they were not obvious features, such as which hand was used or the number of hands used. To help code the above characteristics, the following classifications were used: The gesture was coded as “Pantomime” if the participant used an open or closed hand that mimicked grabbing or moving the object, but the gesture was coded as “Body-as-Object” if a pointing gesture was used. For example, if the participant closed his fist and moved his hand, this was interpreted as a pantomime. If the participant used a pointed finger and moved his hand, this was coded as “Body-as-Object” because the shape of the hand was not pantomiming the physical action of moving an object[5]. Next, a gesture was coded as “Dynamic” if it involved movement, and all gestures involved movement so there were not gestures coded as “Static.” Then, a “Main Action” was considered any gesture representing the intended movement of the action by the hands (e.g., using the hand to move the object), as opposed to a hand or hands used as a peripheral or supporting action (e.g., using the hand to hold an object). All of the actions depicted the intended movement of the object, so all of the gestures were coded as “Main Action.”

---

[4] Grandhi et al. (2011) list one more characteristic that was not used in the current study, “Whether or not tool or object was gestured.” This characteristic was not applicable because the object being acted upon was not specified in the directions to participants to avoid influencing the participant and allow for actions generalizable to computer commands in the current system.

[5] This distinction was made based on the literature indicating that pointing is a communicative referent (i.e., deictic gesture) that is developed very young and well before children begin to switch from body-as-object gestures to pantomimic gestures (Weidinger, Lindner, Hogrefe, Ziegler, & Goldenberg, 2017).

Two coders rated the videos of participants gesturing for each action. The agreement between coders was 100%, so statistical measures of interrater reliability (e.g., kappa) were not calculated. The complete agreement between coders highlights the extent to which the gestures were distinct and understandable as representing the intended action.

A majority of participants used the same hand or hands for each gesture. Between 71%-94% used their right hand only for the actions of moving an object up, down, left, right, clockwise, counter-clockwise, and select. Both hands were predominately used for the actions enlarge and shrink. All but one of the actions were pantomimed gestures of the object manipulations, and the remaining action was selecting an object, of which 76% of participants used a pointing gesture. Every participant performed the gestures dynamically for each action, and all gestures represented the main action of object manipulation. I extended the work by Grandhi et al. (2011) by classifying additional features not included in the above list. Gesture features were also recorded, such as the direction of movement and shape of the hand, to narrow down the converging features of a natural gesture for each action to create the most natural gesture-based computer commands (Table 5). Again, the most often gestured features of each characteristic are shown in bold. The first 10% of gestures were coded by both of the same raters as above. The coding again matched completely, so the remainder of the gestures was coded by one coder.

By recording features not coded in the scheme outlined by Grandhi et al. (2011), it was apparent that the gesture features for each action had limited variation, with many converging features. For each action, all gestures involved movement in the direction of the intended action. For example, every gesture for “move an object up” was performed in an upward motion. Likewise, all participants performed the enlarging gesture by moving hands in an outward-from-center direction, and the same expected motion followed for the other actions except one; only the select gesture had variation on movement direction, with all but one participant pointing or pressing the hand forward while the remaining participant closed their hand in a grasping motion. Also, gestures were all performed within the space of the torso area.

Variation in the gestures occurred mostly in the detailed movements. For the clockwise and counterclockwise rotating actions, gestures varied by the fulcrum point of the rotation, such that a slight majority of participants rotated the hand from the wrist and the rest rotated the entire arm from the elbow. The shrinking and enlarging gestures were all performed in the expected inward and outward directions, respectively, but they varied in whether they moved left-and-right, up-and-down, or diagonally. Although there were variations in the details of movements, these variations were limited to three or less distinctions for each action, and all were performed in the overall expected direction of movement, so defining converging features for each action started from a small set of variations.

Table 5. Converging Features of Gesture Produced for Each Action

	Gesture Features				
Action	**Hand Shape**	**Direction**	**Start**	**Stop**	**Detailed Movements**
1. Up	**70.59% Open** 5.88% Closed 23.53% Point	100% Up	100% Torso Height	100% At/Above Head	N/A
2. Down	**64.71% Open** 5.88% Closed 29.41% Point	100% Down	100% At/Above Head	100% Torso Height	N/A
3. Left	**52.94% Open** 11.76% Closed 35.29% Point	100% Left	100% Torso Height	100% Torso Height	N/A
4. Right	**52.94% Open** 5.88% Closed 35.29% Point	100% Right	100% Torso Height	100% Torso Height	N/A
5. Clockwise	**64.71% Open** 0% Closed 35.29% Point	100% Clockwise	100% Torso Height	100% Torso Height	**64.71% Rotate Wrist** 35.29% Rotate Arm
6. Counter-clockwise	**64.71% Open** 0% Closed 35.29% Point	100% Clockwise	100% Torso Height	100% Torso Height	**64.71% Rotate Wrist** 35.29% Rotate Arm
7. Select	11.76% Open 11.76% Closed **76.47% Point**	**94.12% Forward** 5.88% Closing	100% Torso Height	100% Torso Height	**88.24% Single Point** 5.88% Double Point 5.88% Closing Hand In
8. Enlarge	**70.59% Open** 11.76% Closed 17.65% Point	100% Outward	100% Torso Height	100% Torso Height	**76.47% Left/Right** 5.88% Up/Down 17.65% Diagonal
9. Shrink	**52.94% Open** 35.29% Closed 11.76% Point	100% Inward	100% Torso Height	100% Torso Height	**70.59% Left/Right** 11.76% Up/Down 17.65% Diagonal

Note: The most often gestured features of each characteristic are shown in bold.

Discussion

This first study aimed to narrow down what natural gestures are produced by participants for common object manipulations that may be implemented in a natural user interface. Participants performed gestures for nine object manipulation actions to narrow down the scope of natural gestures to be used as gesture-based computer commands. The majority of participants performed very similar gestures for each action with limited variations, making it possible to define the converging features of gestures for each action. For example, a majority of participants usually used their right hand to perform these representational gestures. This is consistent with previous research; right-handed individuals have shown a preference for using their dominant hand for gestures that represent objects or spatial relationships, but do not have this hand preference for non-representational gestures (Sousa-Poza, Rohrberg, & Mercure, 1979). To confirm whether the most commonly performed gestures in Study 1 are perceived as natural, a second study was conducted in which a separate sample of participants rated these and other gestures along on a scale from natural to arbitrary. Table 6 describes the top natural gestures performed in Study 1 that were included in Study 2 to further narrow down the natural gestures for the experimental testbed.

There are limitations of Study 1 that reduce the generalizability of results beyond the use of developing gestures for the current testbed. One limitation was that the sample of participants was young adult students and naturally produced gestures for this group may not be natural for others, such as older adults. The sample was also right-handed, due to technical limitations of coding the gesture sets. The gestures outlined in Study 1 serve the purpose of creating a starting point for natural gestures to be used in a specific computer lesson, so future natural user interfaces that may implement these gestures should confirm the gestures are perceived as natural for that particular interface.

Table 6. Description of Most Commonly Produced Natural Gestures for Each Action

Action	Description
1. Up	Raise open right hand from chest height to above head with palm forward
2. Down	Lower open right hand from above head to chest height with palm forward
3. Left	Move open right hand from right to left at chest height with palm forward Move pointing right hand from right to left at chest height
4. Right	Move open right hand from left to right at chest height with palm forward Move pointing right hand from left to right at chest height
5. Clockwise	Rotate open right hand clockwise at chest height, circling from elbow Rotate open right hand clockwise at chest height, circling from wrist
6. Counter-clockwise	Rotate open right hand counter-clockwise at chest height, circling from elbow Rotate open right hand counter-clockwise at chest height, circling from wrist
7. Select	Point forward once with right hand at chest height Grasp with right hand at chest height
8. Enlarge	Move open hands outward left and right from center of chest Move closed hands outward left and right from center of chest
9. Shrink	Move open hands inward left and right to center of chest Move closed hands inward left and right to center of chest

CHAPTER FIVE: STUDY TWO

How Are Gestures Interpreted?

The goal of the second study was to assess quantitatively how natural or arbitrary the potential gesture interactions were with a broader range of participants. Natural gestures were determined from Study 1, and potential arbitrary gestures were chosen from a selection of gestures that do not pantomime a real-world physical action from the motion tracker's software development kit (i.e., pre-existing gestures commands recognized by the Kinect) and gesture-based commands from previous experiments (Schroeder, Bailey, Johnson, & Gonzalez-Holland, 2017). In Study 2, participants were asked to rate how natural or arbitrary a gesture seemed for a particular interaction. For each combination of gesture (e.g., moving a hand up with palm facing forward) and desired computer action (e.g., moving an object up), participants were shown a video of an actor performing a gesture and asked what happened in each video and to rate the naturalness of each gesture-action combination. The videos included the natural gestures from Study 1 as well as arbitrary gestures chosen as a comparison. It was expected that for each computer action, the gesture(s) rated as most natural would be the gestures produced for that action in the first study. It was predicted that the arbitrarily chosen gestures would be rated as arbitrary for each action, and gestures from Study 1 that did not match a computer action would also be rated as arbitrary for that combination of action and gesture (e.g., hand moving up gesture from Study 1 for the "select an object" computer action). The results from Study 2 showed whether the gestures produced in Study 1 were also perceived as natural when rated by others.

Method

Participants

A new sample of 188 participants from the same university completed Study 2, and they were awarded class credit. Participants were removed from analyses (*n*=19) if they: 1. Did not respond to all questions, 2. Completed the study in under 10 minutes, 3. Did not describe a video accurately when asked "What happened in the video" (e.g., "no clue"), and/or 4. Did not complete the survey on a computer as required by the instructions (i.e., used a mobile device). The following analyses included 169 participants. Participants were 68% female (*n*=115) and 31% male (*n*=53), and 1 participant chose the answer, "prefer not to respond." Participants were between 18-41 years old with an average age of 20.46 (*SD*=3.90). Participants reported their ethnicity and were able to select multiple options to describe themselves: 54% were White (Non-Hispanic), 24% were Hispanic/Latino, 12% were Asian/Pacific Islander, 8% were African-American, 3% were Arabian/Middle Eastern, 3% selected "Other," 0% were Native American, and 1% chose not to respond (results were rounded to the nearest percent). A majority of participants were right-handed (*n*=146, 86%), with seventeen left-handed participants (10%) and six ambidextrous participants (4%). Mean ratings did not differ significantly by handedness, so responses were collapsed for all analyses.

Materials

Gesture Videos

Twenty-six gesture videos (3-5 seconds each) were presented in which an adult male actor performed each gesture. Eighteen videos corresponded with the natural gestures determined from Study 1 and eight arbitrary gestures were included for comparison (Table 7).

Table 7. Description of Gesture Videos

Natural Gestures Defined from Study 1	1. Raise hand from chest height to head height with open right palm facing up 2. Raise hand from chest height to head height with open right palm forward 3. Lower hand from head height to waist with open right palm facing down 4. Lower hand from head height to waist with open right palm facing forward 5. From chest height, move right hand from right to left with open right palm forward 6. From chest height, move right hand from right to left with right hand pointing 7. From chest height, move right hand from left to right with open right palm forward 8. From chest height, move hand from left to right with right hand pointing 9. From chest height, move right arm with open palm forward making a circle clockwise 10. From chest height, move right hand with open palm forward to make a wrist rotation clockwise 11. From chest height, move right arm with open palm forward to make a full circle counterclockwise 12. From chest height, move right hand with open palm forward to make a wrist rotation counterclockwise 13. Point forward once with right hand at chest height 14. Grasp with right hand at chest height 15. Use both open hands to move outward from center of chest 16. Use both closed fists to move outward (left and right) from center of chest 17. Use both open hands to move inward to center of chest starting with hands about two feet apart 18. Use both hands with closed fists to move inward to center of chest starting with hands about two feet apart
Arbitrary Gestures	19. Bring right arm up to head height making a 90-degree angle at the elbow 20. Move right arm straight down past hip 21. Extend straight right arm out to the right side, parallel to the ground 22. Raise right closed fist to left shoulder 23. From chest height, move from right to left with closed right fist 24. From chest height, move from left to right with closed right fist 25. Use right closed fist to press forward 26. Right open hand wave with palm forward

Demographics

The demographic survey (Appendix A) consisted of 18 items asking for participants' age, sex, ethnicity, etc., as well as educational information. Additionally, questions were adapted from the Video Game Experience Questionnaire (Newcombe & Terlecki, 2005) and included in the demographic survey. The Video Game Experience questionnaire asked questions such as how frequently participants played video games, how many hours participants played a week, what genre of video games participants played (e.g., sports, first-person shooter, puzzle, etc.), and on what device participants played (e.g., console, PC, phone). This demographic questionnaire was also used in the main experiment.

Procedure

Study 2 was conducted entirely online, and participants could respond to questions at their own pace. Participants were recruited from the university's research participation system and were directed to the Qualtrics survey website. After reading an informed consent and agreeing to participate, participants completed the demographic questionnaire. Participants were then told they would be viewing a series of videos depicting gesture-based commands for computer actions, and their task was to rate how natural or arbitrary each gesture was for each command. To reduce ambiguity in the interpretation of the terms "natural" and "arbitrary," participants were told that, "*Natural* means that a gesture is more intuitive or 'makes sense' for that computer action" and that, "*Arbitrary* means that a gesture seems random or doesn't 'make sense' for that computer action." Next, participants were directed to view each gesture video successively and presented in a random order. For each gesture video, participants watched the video and were asked to describe what happened in each video in a text box. After describing the gesture video, the participant then rated the gesture on a 6-point Likert-type scale with endpoints "Completely Arbitrary" and "Completely Natural." Afterward, participants received a post participation debriefing and were awarded credit for their participation. The median time to complete the survey was 23 minutes. Participants could take as much time as they desired to submit this online survey, including starting the survey and returning later; therefore, the range in response times varied widely, from 8 minutes to 2.5 days.

Results

The design was not fully combinatorial – not every action was rated for each of the 26 videos – because opposite gesture-action combinations were not included. For example, a gesture motioning upward was not included in possible ratings for the action of moving an object down. This was done so as not to confuse participants with opposite gesture-action combinations (i.e., “trick questions”), and because opposite movements from actions may represent another separate category of gestures (i.e., opposite actions) and should not be rated on a continuum of natural to arbitrary. For each computer action, a repeated-measures ANOVA was conducted to determine whether there were differences in naturalness ratings of the gestures. Unless otherwise noted, the Greenhouse-Geisser correction is reported for ANOVA tests and Bonferroni corrections were performed for post hoc tests.

Natural Gestures

The top five most natural gesture videos for each action are shown in Figure 3, illustrating that the gestures rated as most natural were those from Study 1, while the remainder gestures were usually rated distinctly less natural than those produced by participants in Study 1.

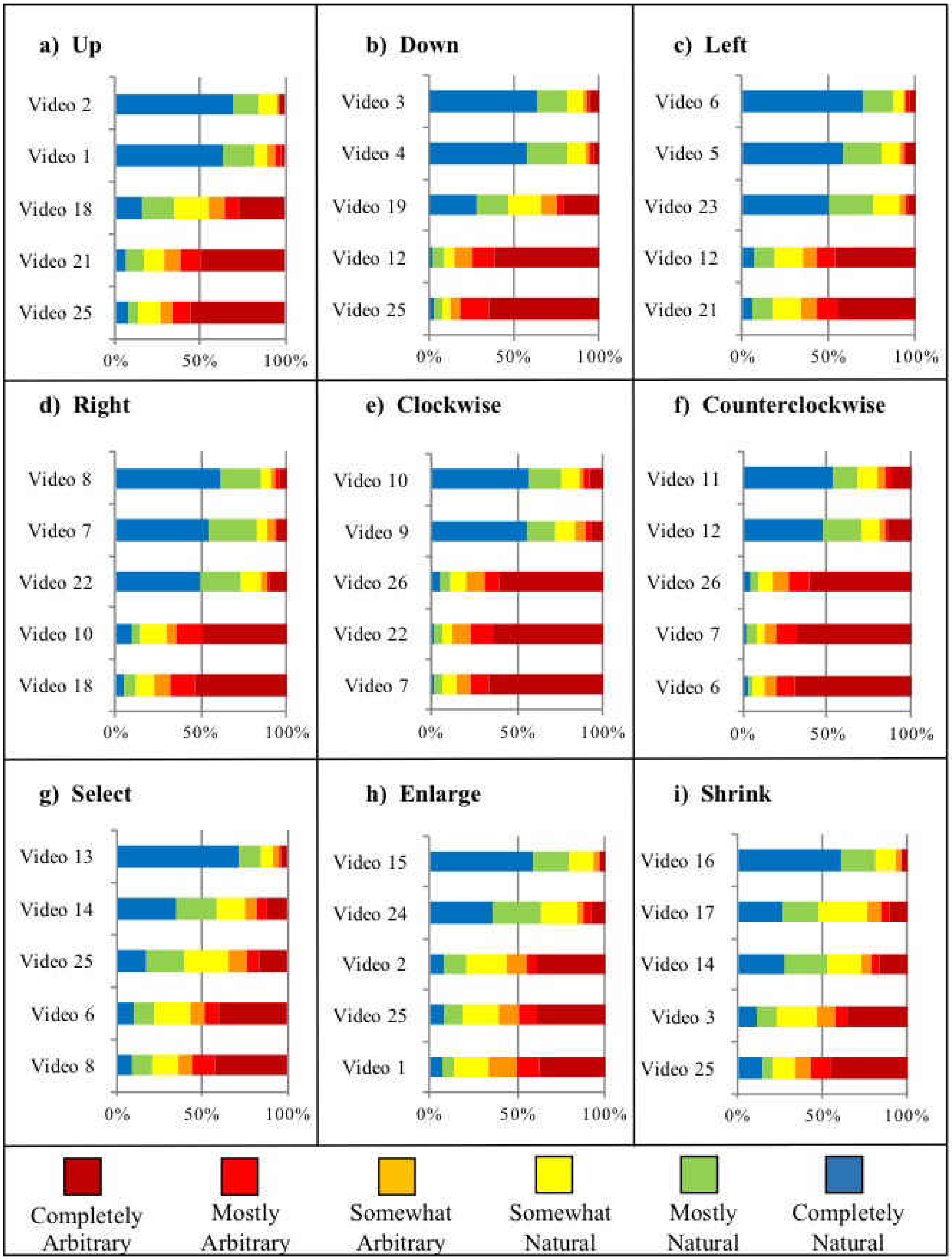


Figure 3. For each gesture-based computer action, gesture videos rated most natural are graphed. Videos higher on the y-axis are seen as most natural for each action. The percentage of responses is presented on the x-axis.

Moving an Object Up

As expected, the naturalness ratings for the action “moving an object up” differed significantly, $F(10.10, 1495.32)=155.34$, $p<.001$, $\eta p^2=.512$. Depicted in Figure 3a, the gestures rated most natural for the action “moving an object up” were Video 2 ($M=5.40$, $SD=1.17$) and Video 1 ($M=5.24$, $SD=1.28$). These two gestures were rated significantly more natural than all of the other gestures (all ps<.001), but did not differ significantly from each other ($p=.12$). The gestures shown in these two videos were those that were produced for “moving an object up” in Study 1.

Moving an Object Down

Naturalness ratings differed among gesture videos for “moving an object down,” $F(7.88, 1213.17)=234.27$, $p<.001$, $\eta p^2=.603$. Figure 3b shows that the most natural gestures were Video 3 ($M=5.24$, $SD=1.32$) and Video 4 ($M=5.23$, $SD=1.22$), which were significantly more natural than the other gestures and did not differ from each other ($p=.11$). These videos were the gestures produced in Study 1. Although the arbitrary gesture depicted in Video 19 ($M=3.95$, $SD=1.85$) in which the arm moves down past the hip was less natural than Videos 3 and 4, it was rated significantly more natural than the other gestures for this action (all ps<.001). This arbitrarily chosen gesture was likely rated more natural than the other gesture-action combinations because it looks similar to a pantomimed gestures – that is, the gesture in Video 19 looked like the pantomimed gestures in Videos 3 and 4, which were natural gestures produced in Study 1; however, it is important to note that Video 19, while similar to pantomimed gestures, was rated less natural than either of the gestures from Study 1.

Moving an Object Left

Gesture ratings were different for the action "moving an object left," $F(7.11, 1095.51)=234.47$, $p<.001$, $\eta p^2=.604$. The gestures rated more natural for this action (all ps<.001) were Video 6 (M=5.40, SD=1.20) and Video 5 (M=5.17, SD=1.34), which were not different from each other (p=.10) and were both from Study 1 (Figure 3c). The naturalness rating for the arbitrary gesture in which a closed fist was moved right to left in Video 23 (M=5.06, SD=1.29) did not differ from the rating for Video 5, but was rated less natural than Video 6. Even though this gesture was arbitrarily chosen, it did resemble the natural gestures in Videos 5 and 6 because it made the same motion with a closed fist as opposed to an open or pointing hand.

Moving an Object Right

The gesture ratings also differed for "moving an object right," $F(8.26, 1304.63)=219.30$, $p<.001$, $\eta p^2=.581$. Shown in Figure 3d, gestures rated most natural (all ps<.001) for the action "moving an object right" were Video 8 (M=5.26, SD=1.27) and Video 7 (M=5.14, SD=1.33) from Study 1, and arbitrary Video 22 (M=4.89, SD=1.56) showing a closed fist moving right. Videos 7 and 8 did not differ at all (p=1.00), and ratings for Videos 7 and 22 (p=1.00) and Videos 8 and 22 (p=.67) were not statistically different. Comparable to the finding for "moving an object left," the arbitrary gesture that was rated closely to the natural gestures from Study 1 for "moving an object right" was very similar to the natural gestures.

Rotating an Object Clockwise

For "rotating an object clockwise," gesture ratings differed, $F(9.81, 1510.63)=152.78$, $p<.001$, $\eta p^2=.498$, with Video 10 (M=5.26, SD=1.27) and Video 9 (M=5.14, SD=1.33) rated significantly more natural than other gestures (all *p*s<.001, Figure 3e). Both gestures were from Study 1 and did not differ from each other (p=1.00).

Rotating an Object Counterclockwise

Ratings of naturalness differed for "rotating an object counterclockwise," $F(7.40, 1132.68)=143.80$, $p<.001$, $\eta p^2=.485$. The gestures rated most natural for the action were from Study 1 (all *p*s<.001, Figure 3f), Video 11 (M=4.77, SD=1.72) and Video 12 (M=4.71, SD=1.72). These did not differ from each other (p=1.00).

Selecting an Object

There was a difference in gesture ratings for "selecting an object," $F(15.44, 2408.51)=71.68$, $p<.001$, $\eta p^2=.315$. Shown in Figure 3g, the naturally rated gestures for the action "selecting an object" were Video 13 (M=5.38, SD=1.21) and Video 14 (M=4.38, SD=1.70). These videos from Study 1 were rated more natural than all other gestures (all *p*s<.001), with Video 13 rated more natural than Video 14 ($p<.001$).

Enlarging an Object

Differences between the gesture ratings were also significant for "enlarging an object," $F(15.35, 2409.90)=81.20$, $p<.001$, $\eta p^2=.341$. Video 15 (M=5.24, SD=1.17) and Video 24 (M=4.62, SD=1.52) were rated more natural than the other gestures (all *p*s<.001), and Video 15 was significantly more natural than Video 24 (p=.003, Figure 3h). Video 15 and Video 24 were gestures from Study 1.

Shrinking an Object

Finally, the gesture ratings for "shrinking an object" were also different, $F(13.40, 2130.80)=104.96$, $p<.001$, $\eta p^2=.401$. The gestures rated most natural for the action "shrinking an object" were Video 16 (M=5.31, SD=1.11) and Video 17 (M=4.34, SD=1.49), but Video 16 was rated significantly more natural than Video 17 (p<.001; Figure 3i). Both Video 16 and Video 17 were gestures produced in Study 1.

Discussion

For each action, the expected gestures (i.e., gestures produced in the Study 1) were rated as most natural. This indicated that gestures people produce when asked how they would naturally perform a gesture-based computer action are interpreted as the intended action by a separate sample of participants. For some actions, arbitrary gestures that were similar to natural gestures were also rated as natural (above the midpoint on a continuum of arbitrary to natural), confirming that gestures which resemble pantomimed actions are considered more natural. This answered the question of which natural gestures to implement as gesture-based commands in the computer lesson by determining how people gesture actions, and how those actions are interpreted.

Selection of Experiment Gestures

Natural Gestures

Based on the results showing which gestures were rated as most natural for each action, the gesture-based commands for the computer interface were determined. The natural gestures for the computer commands were chosen from the top-rated natural video for each action. If more than one gesture was rated as most natural for an action, the gesture that more closely resembled the gestures for other actions was chosen. For example, the action of "moving an object up" had two gestures rated as most natural in which the right hand was moved from the torso to above the head – one with the palm facing forward and the other with the palm facing upward. Although these gestures did not significantly differ from each other on ratings of naturalness, the gesture in which the palm was facing forward was chosen for the computer command because the palm facing forward is a similar feature with other naturally-rated gestures. Figure 4 depicts an actor performing each of the top-rated natural gestures that were included in the computer lesson testbed for each action.

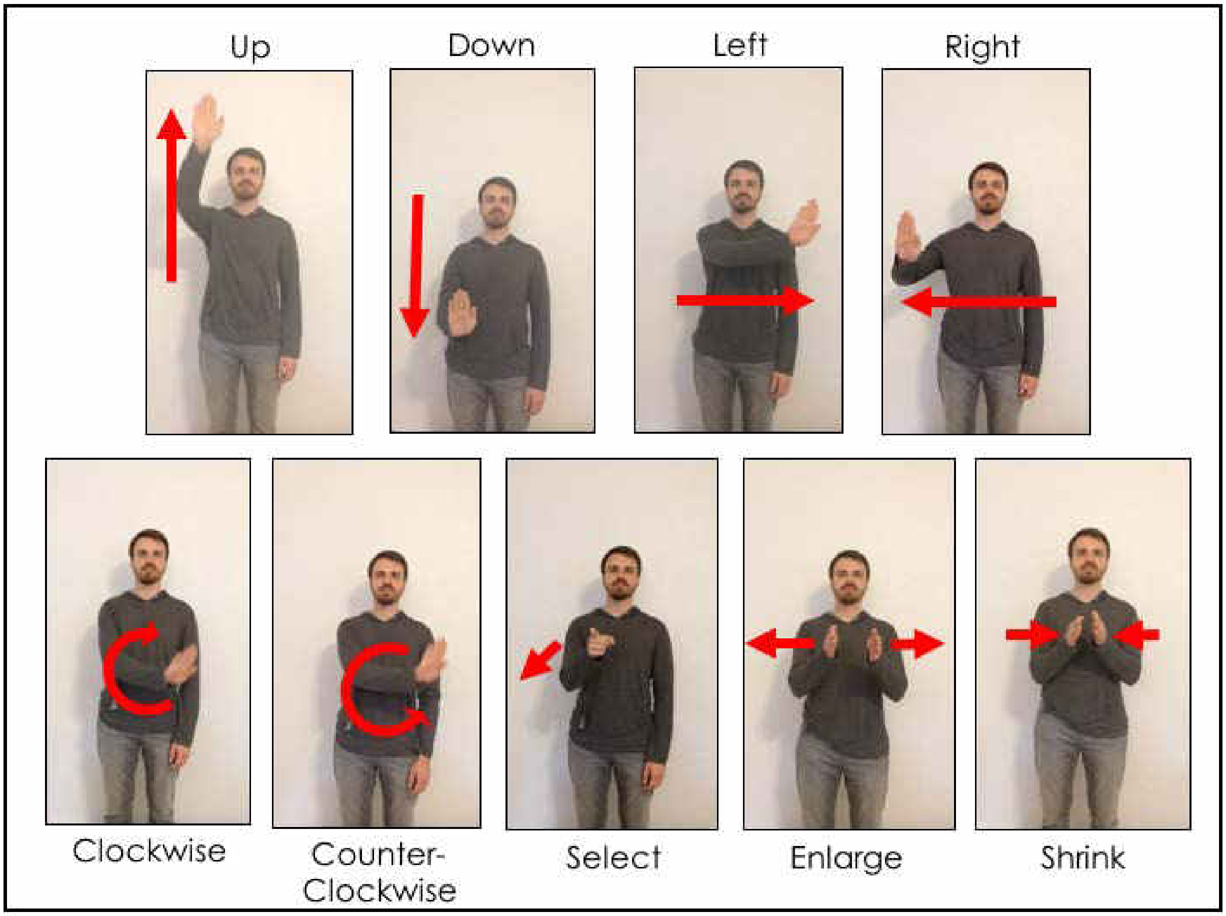


Figure 4. Natural Gestures for Computer Commands

<u>Arbitrary Gestures</u>

For each action, the remaining videos were rated on the arbitrary end of the scale between "Mostly Arbitrary" and "Completely Arbitrary," and there was little difference in the extent to which the arbitrary gestures were rated. To choose the arbitrary gestures, I looked at every gesture rated between "Mostly Arbitrary" and "Completely Arbitrary" that was not rated significantly less arbitrary than any other gesture. It is important to note again that none of the arbitrary gesture-action combinations were opposite actions from the action, such as gesturing leftward for the “moving and object right” action. To narrow down which arbitrary gesture to select for each action from the equivalently arbitrary gestures, I first used gestures with similar features as those found in the set of natural gestures. For example, the “enlarge” gesture for the natural gesture condition (i.e., open palms are moved outward from the center of the body) was rated as arbitrary for the action of “select,” so this gesture was used in both the natural and arbitrary conditions but for different actions. Next, for complementary gestures (e.g., up and down; right and left), I found two arbitrary gestures that were the opposite of each other, such as fists moving inward for the “up” action and fists moving outward for the “down” action. Because the natural gestures often used complementary gestures for corresponding actions (i.e., moving the hand up/down for “up” and “down” actions), I chose to use complementary gestures for corresponding actions in the arbitrary gesture condition because this allowed chunking of gestures in a similar way to the chunking of natural gestures. This prevented the set of arbitrary gestures from consisting of nine distinct gestures while the natural set of gestures consisted of several pairs of gestures. If complementary gestures did not exist in the remaining set of arbitrary gestures, I chose the gestures that could be chunked by similar features. For example, the arbitrary clockwise gesture was a forward grasping motion and the counter-clockwise gesture was a grasp with arm moving to the side (Figure 5). Additionally, each gesture was tested using the Microsoft Kinect V1 motion-capture linked to the Unity 3D game engine to confirm that these gestures were capable of being recognized by the computer system. Every gesture from both sets was able to be recognized by the motion tracker.

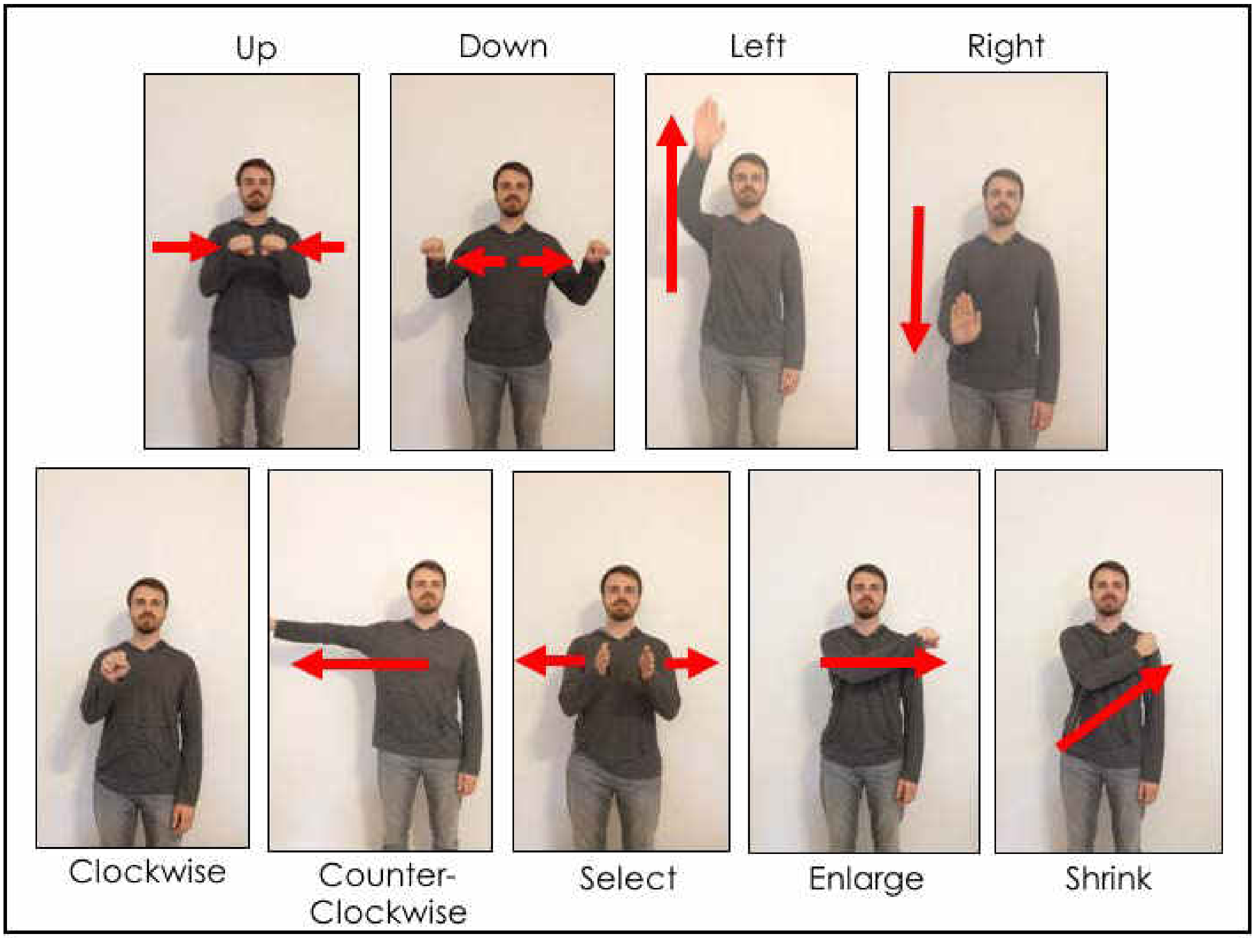


Figure 5. Arbitrary Gestures for Computer Commands

CHAPTER SIX: EXPERIMENT

The main experiment manipulated the type of gesture-based interaction and the instructions of those gestures to determine whether these instructional techniques differ for learning conceptual information in a computer lesson. The experiment was used to answer the research questions of whether natural gestural interactions result in more learning than arbitrarily-mapped gestures, and whether video or text-based instructions for the gesture interactions can influence the computer lesson. The computer lesson involved conceptual information on optics, and participants learned how lenses and mirrors interact with beams of light by manipulating lenses and mirrors using the gesture-based computer interactions. The instructional techniques were assessed by comparing the amount of conceptual information learned and the amount of mental effort required to complete the lesson. Prior to participating in the in-lab experiment, participants completed an online prescreening.

Prescreening

All prescreening measures and questionnaires were completed online prior to the in-lab experiment. Participants who completed the entire prescreening as indicated by the research participation system were sent requests through the research system to participate in the experiment.

Prescreening Participants

Three hundred people completed the prescreening study online. Of the participants who completed the study, 128 participated in the main experiment (described in the Experiment Participants subsection below). Participants were excluded from signing up for the prescreening if they were younger than 18 years old or were not predominately right-handed.

Materials

Knowledge of Optics Pre- and Post-Test

Participants' knowledge of optics was measured in a pre-test during the prescreening portion of the experiment that occurred online prior to the in-lab experiment. The purpose of the pre-test was two-fold: First, the pre-test was used to screen participants so that those with more incoming knowledge of optics were not included in the final analyses comparing pre- and post-test scores. Second, the pre-test was used to assess how much information was gained from the computer lesson by comparing the change score between the pre- and post-test scores (Δ).

The Knowledge of Optics Test (Appendix B) was developed by adapting questions from middle school (ages 11-14) science textbook test banks (*Science Voyages*, 1999) and online physics lesson resources developed by Florida State University, University of Florida, Los Alamos National Laboratory, and The Optical Society (Davidson, 2015a, 2015b, 2015c; Henderson, 1999a, 1999b). To determine a participant's conceptual understanding of optics, the test included 29 fill-in-the-blank and multiple choice items asking how light reflects and refracts, types of lenses and mirrors, and applying that information to different lens and mirror placements. All of the test questions could be answered by recalling and applying the information presented in the computer lesson (see Appendix C for screenshots of the computer lesson).

Paper Folding Test

Spatial ability was measured using the Paper Folding Test (PFT; Ekstrom, French, Harman, & Dermen, 1976). The PFT is a 10-item timed measure of spatial visualization. The test asks participants to imagine what a piece of paper would look like if it was folded and then a hole was punched through the paper. The task was to select from five possible answer choices what the paper would look like when it was unfolded. Participants were given three minutes to complete the test. The maximum score possible was 10 points. Participants were awarded one point when a question was answered correctly. Participants lost one-fifth of a point for each incorrect answer to discourage random guessing. If a participant did not respond to a question, no points were awarded or lost.

Brief Assessment of Gesture Survey

The Brief Assessment of Gesture survey (BAG; Nagels et al., 2015) is a 12-item measure of attitudes toward gesturing and gesturing behaviors that was included as an individual difference measure that may affect performance on the gesturing task (Appendix D). The BAG is divided into four factors: Perception, Production, Social Production, and Social Perception. The first factor, Perception, measures the extent to which participants perceive gesturing in an unfavorable way, such as “I find it very annoying when I'm talking to someone who gestures a lot during the conversation.” A higher score on the Perception factor indicates a negative perception of gesturing. The second factor is Production, or the propensity of the participant to produce gestures in communication, and the degree to which participants enjoy others gesturing. An example item in the Production factor is, “I've been told before that I gesture a lot when I talk.” The next factor is Social Production, measuring the participant’s use of gestures in goal-oriented communication, including, “When talking in noisy places, I usually gesture a lot to make myself understood over the noise.” The final factor was Social Perception, or the extent to which participants were surprised or amazed at others gesturing; “I often feel amazed by people who are able to gesture a lot when they talk.”

Demographics and Video Game Experience

The demographic survey used in the main experiment was the same as that used in Study 2 (Appendix A).

Attention Check Questions

Several attention check questions were included for use in removing participants who did not carefully read the questions. Before starting the questionnaires, participants must have answered "Yes" to the statement, "I will answer all questions honestly and to the best of my ability." In the BAG section, participants should have answered the statement, "I did not pay attention to the questions in this study" with the response "Not Agree." In the Knowledge of Optics Pre-test, the question was asked, "Are you reading all the questions and answering honestly?" The response format for this question was fill-in-the-blank, so participants must have indicated an affirmative answer to be considered for the main experiment (e.g., "Yes," "to the best of my ability," "yes im trying here [sic]," etc.). For details on how these questions were used to remove participants and the number of participants removed, see Participant Removal subsection later in this chapter.

Prescreen Procedure

Participants signed up for the experiment through the university research system website and were directed to a Qualtrics link where they read an informed consent page. The questionnaires and measures were completed online. The order of tasks was randomized for each participant. To confirm that participants completed each questionnaire without skipping questions, participants were instructed not to leave any questions blank, and to write "I don't know" if they did not know the answer to a question. Once the tasks were completed, participants were given a post participation form explaining the purpose of the prescreening was to determine participants for the main experiment, and they may be contacted to participate in a future study.

Experiment

Participants who completed the online prescreening were invited to participate in the in-lab experiment. All participants who completed the prescreening were contacted through the university's research participation system to sign up for experiment participation.

Design

To parse out the effects of natural mapping and instruction, participants were randomly assigned to one of four conditions resulting from fully crossing the two levels of both factors (2X2 between-subjects design). The independent and dependent variables included in the experiment are listed in Table 7.

Table 7. Table of Independent and Dependent Variables

Variable	*Manipulation*	
IV1	A.	Video instructions
	B.	Text instructions
IV2	A.	Natural gestures
	B.	Arbitrary gestures
Variable	*Measurement*	
DV1	Knowledge of Optics Δ score	
DV2	Cognitive Load Questionnaire score	
DV3	Presence Questionnaire score	
DV4	System Usability Questionnaire score	

Conditions

The four between-subjects conditions in this experiment are a result of crossing the two levels of both independent variables shown in Table 7. The table below contains a detailed explanation of each condition as it was implemented in the experiment (Table 8).

Table 8. Description of Conditions in the Experimental Task

Condition	*Condition Description*
Condition 1	**Naturally-mapped gesturing with video instructions:** During the tutorial instructions before the experimental task, participants were instructed on the gestures they will use in the testbed by watching video instructions of an actor performing the gestures. The gestures used to interact with the computer lesson were the most natural gesture-based commands determined in Studies 1 and 2.
Condition 2	**Naturally-mapped gesturing with text instructions:** The tutorial instructions included short, text-based directions on how to perform each gesture. The gestures were the same natural gestures used in Condition 1 that were determined from Studies 1 and 2.
Condition 3	**Arbitrary gesturing with video instructions:** The video instructions in the tutorial depicted the same actor from the video tutorial in Condition 1 performing the arbitrary gestures. The arbitrary gestures were those rated as most arbitrary for each action in Study 2.
Condition 4	**Arbitrary gesturing with text instructions:** Text-based instructions for the arbitrary gestures were presented in the tutorial. The arbitrary gestures were the same arbitrary gestures presented in Condition 3 that were determined from Study 2.

Materials

Tutorial

Tutorials for each condition were presented to participants on Microsoft Powerpoint before starting the computer lesson. The tutorial explained the gestures using either video- or text-based instructions. The tutorials explained that participants would be using gestures learned in the tutorial to complete a computer lesson. Then, each of the nine gestures for the computer actions (e.g., move an object up, select an object, etc.) were presented with one gesture per slide. There were 15 slides total for each tutorial. Participants proceeded to the next slide by clicking the mouse. Gestures were presented in the same order for each condition. Participants could complete the tutorial at their own pace and review slides as desired. Participants completed the slides in approximately 10 minutes, although time to complete the tutorial was not measured.

After the gestures were presented, a slide instructed participants to recall the gestures in the same order in which they were learned. On the final slide, all participants were instructed to perform the gestures in a random order (the same random order was given in each condition) so every participant had the opportunity to learn the gesture three times before completing the tutorial. When the participants performed the gestures at the tutorial, the experimenter watched to confirm the gestures were accurate before proceeding to the computer lesson. If the gestures were incorrect, the experimenter instructed the participant to review the slide for that gesture and answered any clarifying questions the participants had. Experimenters were explicitly instructed not to show participants how to do the gestures by physically performing the gestures.

Experimental Testbed

The experimental testbed was a computer lesson called "Hubble Needs Glasses," which teaches how light interacts with mirrors and lenses using gesture-based computer commands.

The computer lesson was presented on a 30" LCD television screen, and the participant stood on a mark nine feet from the screen while performing the gesture-based computer interactions that were recognized by the motion tracker (Figure 6).

The testbed was developed using the 3D Unity game engine, which presented information in a slide-like format with two interactive sections. There were nine slides in the lesson, of which seven were instructional content and two were the interactive sections (for screenshots of the all the instructional content slides and interactive gesture sections, see Appendix C). In the interactive sections, participants used gestures (either Arbitrary or Natural gestures, depending on condition) to manipulate mirrors and lenses in a beam of light to learn how light interacts with each type of mirror or lens. A narrator read the information presented on the slides and instructed participants on which object manipulations should be performed. The object manipulations were activated via gesture-based computer commands.

The gesture-based interaction commands differed for the Natural and Arbitrary conditions, and the respective gestures were determined from Studies 1 and 2 (see Figures 4 and 5). The gestures were implemented using the Microsoft Kinect V1 motion tracker. During the interactive sections in which participants manipulated mirrors and lenses, the participant was instructed to perform a gesture indicating a specific object manipulation. For example, the narration instructed the participant to, "Select the concave mirror. Move the mirror down into the beam of light and enlarge it." The participant then performed the three gestures sequentially (i.e., "select," "down," and "enlarge"). When each gesture was recognized by the Kinect for the corresponding computer action, an animation was triggered of the sequence of actions. The result of each action was seeing how the beam of light interacted with the mirror or lens being manipulated.

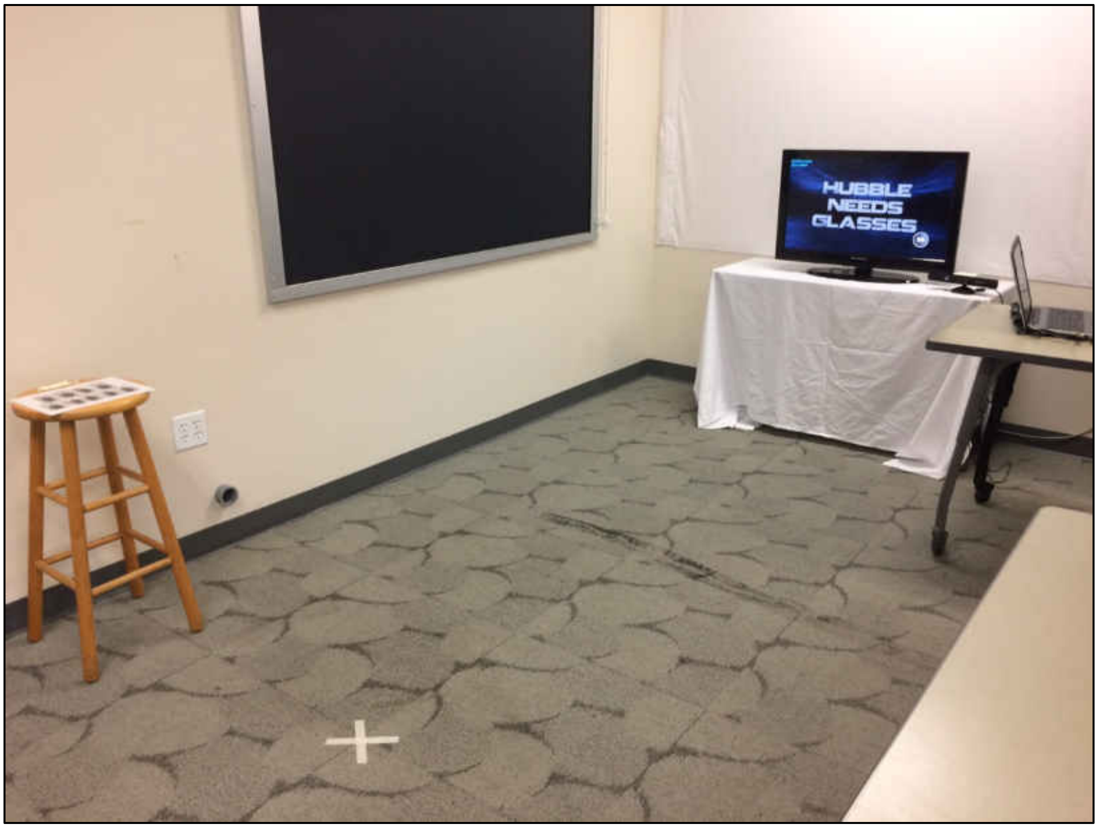


Figure 6. Experiment Room Setup. Participants stood on the "X" and faced toward the television screen that displayed the computer lesson. The motion tracker was positioned next to the television screen. The gesture reference sheet was placed on the stool next to the participant. The experimenter operated the computer lesson from the laptop (next to the television screen).

The computer lesson began with a description of the Hubble Space Telescope, followed by explanations of how mirrors were used to direct light to focus images of the universe. The lesson then explained the concept of refraction and described types of mirrors. Following this section, participants completed the interactive section in which they used gestures to manipulate a mirror in a beam of light to see how light reflects (Figure 7). There were three mirrors in this section: planar, concave, and convex. Each mirror was moved into the beam of light by the participant and the narrator instructed which gesture-based actions to perform. After the participant interacted with each mirror, they advanced the slide to the refraction and lenses section. This section described the concept of refraction as well as converging and diverging lenses. The participant then completed the interactive section with four types of lenses: bi-concave, planar-concave, bi-convex, and planar-convex. When the participant finished manipulating each of the four lenses in the beam of light, the computer lesson ended.

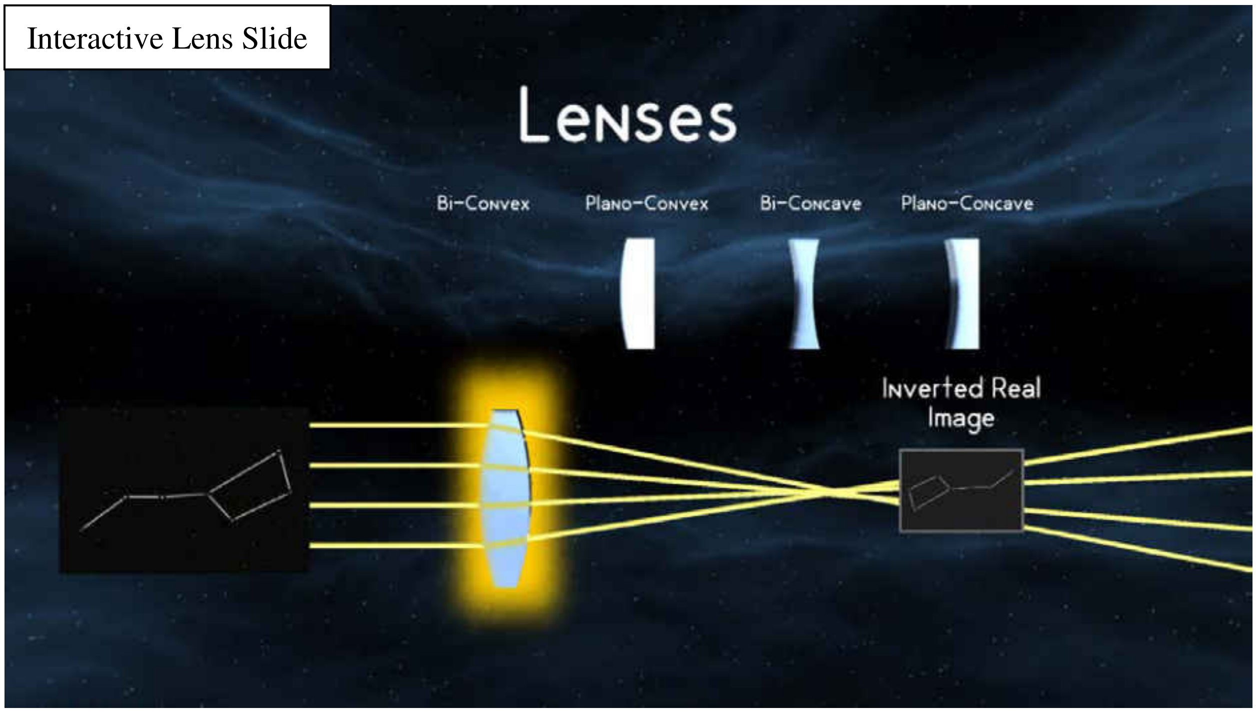


Figure 7. Screenshot of the interactive section for lenses. Participants moved the lens into the beam of light by using the “select” gesture to highlight the lens, then the “down” gesture to move the lens into the beam of light. Next, participants used the “enlarge” gesture to increase the size of the lens. When the lens was moved into the beam of light, the light refracted, illustrating the conceptual information of refraction. The same type of interaction was completed for mirrors, illustrating the concept of reflection.

Gesture Reference Sheets

For each condition, participants were given a gesture reference sheet corresponding to the type of gesture (i.e., Natural or Arbitrary) and method of instruction (i.e., Video or Text). The reference sheets were provided based on pilot testing which suggested some participants would not be able to recall all nine gestures and would be unable to complete the lesson without reminders. The impact of not recalling the gestures would be not being able to complete the computer lesson, and completion of the lesson was not one of the main outcome variables to answer the research questions; therefore, because completion of the lesson was necessary to answer the research questions, reference sheets were provided to assure participants would be able to finish. The reference sheet was placed on a stool next to the participant during the computer lesson so that it could be easily referenced throughout the experiment. The reference sheets for each condition consisted of a single page with all nine gestures (Appendix E). The reference sheets contained either pictures for the video conditions or blocks of text for the text-based conditions, and gestures were either Natural or Arbitrary. For the video instruction conditions, a picture of each gesture was taken from screenshots of the videos, and red arrows were overlaid on the pictures to indicate the direction of movement.

Presence Questionnaire

The Presence Questionnaire (PQ; Witmer, Jerome, & Singer, 1998) contains 19 items in which participants report how much they felt "present" in a training environment. Participants responded on a 7-point Likert-type scale (the anchors differed depending on question, see Appendix F). The PQ was used because the questions used to determine the "sense of being there" in the training environment are also applicable to how much control the participant felt and the naturalness of the interactions; so although the construct of "presence" was not investigated *per se*, the PQ may measure the perceived "naturalness" of interacting with the computer lesson. For example, "How much did your experiences in the virtual environment seem consistent with your real world experiences?" and "How much did the control devices interfere with the performance of assigned tasks or with other activities?" To see all of the questions on the PQ, refer to Appendix F.

Overall sense of presence is measured by averaging the items on the PQ, and the PQ can be divided into four subscales: Involvement, Sensory Fidelity, Adaptation/Immersion, and Interface Quality. The Involvement subscale measures the degree to which the participant feels the control of the computer interface is natural. Sensory Fidelity is the feeling that the senses are engaged in the system and operate as expected (i.e., sounds can be identified and localized). Adaptation/Immersion is the ability of the participant to adapt to the computer environment and concentrate on the activities presented in that environment. The last subscale, Interface Quality, is the extent to which the interaction with the computer task distracts from or otherwise hinders performance in the virtual environment. Scores on each of these subscales may be affected by the naturalness of the gestures the participant is assigned.

Cognitive Load Questionnaire

The Cognitive Load Questionnaire (Paas, van Merriënboer, & Adam, 1994) was chosen because it is a frequently used single-item measure of perceived mental effort. The item asks participants to, "Please indicate on the scale your level of mental effort on the task you just performed. Think only about your level of effort on the task you performed immediately preceding this questionnaire." Paas, Tuovinen, Tabbers, and van Gerven (2003) explain that, "The scale's reliability and sensitivity and moreover its ease of use have made this scale, and variants of it, the most widespread measure of working memory load within CLT [Cognitive Load Theory] research" (p. 68). The scale was a 10-point Likert-type scale with anchors "Very, very low" to "Very, very high," with a higher rating indicating higher perceived mental effort for the computer task.

System Usability Scale

The System Usability Scale (SUS; Brooke, 1996) is a 10-item measure that indicates how usable the system (i.e., computer lesson) seemed to participants. Participants respond to questions on a 5-point Likert-type scale with the endpoints "Strongly Disagree" and "Strongly Agree." Example questions include, "I thought the system was easy to use" and "I found the various functions in this system were well integrated." Higher average SUS scores indicate better perceived usability of the system.

Manipulation and Attention Checks

Several questions were included in the tasks after the experiment to confirm participants viewed the manipulation as intended and participants were paying attention throughout. The manipulation check question presented after the computer lesson portion of the experiment was, "Rate how natural or arbitrary you though the gestures were to interact with the computer system." Participants responded on a 6-point Likert-type scale from "Completely Arbitrary" to "Completely Natural." Another manipulation check to verify participants were paying attention during the gesture tutorial was an open-ended item, "Describe the gesture you used to select an object." Finally, to determine whether participants were paying attention, an open-ended attention check question was asked during the post-test measure, "Are you reading all the questions and answering honestly?"

Experiment Participants

Power Analysis

To determine the appropriate sample size for the 2X2 fixed effects ANOVAs, a power analysis was conducted using G*Power software (Faul, Erdfelder, Lang, & Buchner, 2007). A medium effect size for a F-test was anticipated ($\eta^2 = 0.25$) at a 0.05 alpha level with 80%

power[6]. The numerator has one degree of freedom, and there are four conditions. The total recommended sample size suggested by G*Power was 128 participants, with n=32 participants in each of the four conditions. The total number of participants who completed the study was 128.

Participant Removal

Of these 128 participants, 26 were removed for the reasons described below, and the final sample size included in analyses was 102 participants. Seven participants were removed due to a glitch in the university participation system that resulted in missing pre-test scores. One participant was removed for not completing the experiment, and nine people were removed for incorrect responses to the attention check questions or not following directions on the spatial measure (e.g., selecting more than one answer for a question). Nine participants were removed for scoring above the cutoff of 56.7% correct on the Knowledge of Optics pre-test.

Participants who scored above this cutoff on the Knowledge of Optics pre-test were excluded from analyses because they came into the study with more knowledge of the learning material in the computer lesson, and there may be a treatment by aptitude interaction such that participants who know more about optics may learn from the computer lesson differently than those who know less about optics. This cutoff score was determined by examining the scores on the pre-test and removing participants whose knowledge of optics was beyond that of most participants. Although the range of scores on the pre-test was high, with the lowest score of zero answers correct and the highest score of 89.29% correct, the distribution of scores was skewed left such that the average participant scored around 21.7% correct (SD=17.5%). The skew and kurtosis values were determined by dividing skewness and kurtosis by their respective standard

[6] The anticipated medium effect size and necessary power level were selected based on commonly used parameters when previous literature does not suggest expected effect sizes.

errors (*Skew*=5.88, *Kurtosis*=4.01), and these values exceeded 3.29, indicating skewness (Field, 2013). Because the scores on the pre-test were skewed such that most participants scored on the low end, participants who scored above two standard deviations from the mean (i.e., 21.7+[17.5*2]=56.7% correct) were excluded from analyses to avoid a treatment by aptitude interaction affect. After removing participants who scored above the cutoff, the pre-test scores were normally distributed[7].

Participant Demographics

All participants were undergraduate students at an university. Participants were excluded from participating in the experiment if they were not predominately right-handed or were younger than 18 years old. The average age of participants was 18.69 years old (*SD*=1.54 years), and participants were 68.6% female (*n*=70), 31.4% male (*n*=32). Participants indicated that 76 were Science, Technology, Engineering, and Math (STEM) majors, 22 were Non-STEM majors, and four were undeclared majors. The race/ethnicities of participants were 9.6% Asian/Pacific Islander, 14% Black/African-American, 27.2% Hispanic/Latino, 1.8% Other, 1.8% selected "Prefer not to respond," and 45.6% White/Caucasian[8].

Procedure

After completing the prescreening online, participants were contacted through the university research system to sign up for the main experiment, which was conducted individually in a lab setting. Upon arrival to the lab, participants read an informed consent and agreed to participate. Participants were randomly assigned to one of four conditions: 1. Natural gestures with video instructions, 2. Natural gestures with text instructions, 3. Arbitrary gestures with

---

[7] More details on the Knowledge of Optics pre-test measure after removing high scoring participants are provided in the next Chapter that describes results of the experiment. These descriptives are not included here because they do not pertain to the removal of participants and are more applicable to the next Chapter.
[8] Participants could select multiple responses to better represent their race/ethnicities.

video instructions, or 4. Arbitrary gestures with text instructions. The experimenter then explained the purpose of the study was to determine what type of gestures are best for interacting with a computer lesson for learning conceptual information. Next, participants completed the tutorial for gesture instructions that corresponded to their condition. The tutorial could be completed at the participant's own pace, and participants were allowed to ask any questions they had during the tutorial. The experimenter could answer any questions about the tutorial without physically performing the gestures. When the tutorial was completed, the experimenter confirmed with the participants that they did not have any more questions about the gesture-based commands before proceeding to the computer lesson. After the tutorial, participants were shown the gesture reference sheet for their respective conditions and told they could use this sheet throughout the computer lesson.

Participants were then directed to stand on a taped "X" on the floor, facing toward the television monitor and motion tracker (see room setup in Figure 6). The reference sheet was placed next to the participant on a stool so that it could be accessed easily throughout the computer lesson without requiring the use of hands while gesturing. The participant was instructed to follow the directions of the narrator on each slide to complete the computer lesson. For each gesture instruction, the participant must wait for the narration to end before performing a gesture. The motion tracker would not recognize a gesture while the narrator was speaking to avoid participants skipping key learning material. To move on to the next slide, participants used the "select" gesture on the arrow that appeared at the end of a slide. The experimenter started the computer lesson from an interface on a laptop connected to the television monitor and motion tracker. The participant completed the lesson described in the Testbed section (above).

The computer lesson took approximately 10-15 minutes to complete (see Results section for more detailed time descriptions).

Once the computer lesson portion of the experiment was completed, participants sat back at the computer on which they saw the tutorial to complete the remaining measures. First, participants rated their mental effort on the Cognitive Load Questionnaire. Then, the Knowledge of Optics post-test was completed with the same questions from the pre-test in a randomized order. Finally, participants rated their perceptions of the computer lesson environment on the PQ and SUS. Manipulation and attention check questions were randomly included within the other scales. The entire in-lab experiment took approximately 45 minutes to 1 hour to complete.

CHAPTER SEVEN: RESULTS

The results presented in this chapter are organized in subsections that reflect the Prescreening and Experimental portions of the study. In the Prescreening subsection, the subject variable measures, which quantify potential covariate predictors of the experimental outcome measures (e.g., Knowledge of Optics learning, Cognitive Load), are described and analyzed to provide context for the use of these subject variables in the experimental analyses. Following the Prescreening subsection, the Experimental subsection includes analyses of the main hypotheses regarding Knowledge of Optics learning and cognitive load, with the potential covariates (as determined by the Prescreening analyses) included in the experimental tests. Then, additional analyses pertaining to peripheral questions are presented, such as how the perceived usability of the system differed by conditions, etc. Following these analyses, the empirical results are outlined in tables in the Summary of Results subsection. Finally, qualitative participant reactions to the computer lesson are reported to expand on how the system was perceived for the different conditions.

Results were analyzed using IBM SPSS v20. Significance was reported at the level of α=0.05, unless otherwise specified. Skew and Kurtosis were calculated by dividing each by their respective standard errors, and the standardized values were considered unacceptable over the absolute value of 3.29 (Field, 2013). Levene's Test of Equality of Error Variances is reported when variances were significantly different.

Prescreening Measures

The following analyses of the prescreening measures include the participants in the final sample after participant removal was conducted (see Participant Removal subsection in Chapter 6 for details).

Knowledge of Optics Pre-test

The Knowledge of Optics pre-test was completed by participants online prior to the in-lab experiment. As described in the Participant Removal section in Chapter 6, participants were removed from analyses if they scored above two standard deviations from the mean on the Knowledge of Optics pre-test, or above 56.7% correct. After removing participants (n=9) who scored above the cutoff value, the pre-test scores were distributed normally around the mean of 20.19% correct (*SD*=12.54%, *Skew*=2.56, *Kurtosis*=-1.22), with a low score of zero correct and the highest score of 50% correct. Average scores on the pre-test did not differ significantly among conditions, $F(3, 98)=0.175$, $p=.913$, $\eta p^2=.005$ (see Table 10 in the Experiment Results subsection for pre-test scores compared to post-test scores by condition).

Paper Folding Test

Spatial ability was measured using the PFT during the prescreening because it is likely a predictor of performance on the experiment. Scores on the PFT were between -2 and 10 points, out of a highest possible score of 10 points. Scores were normally distributed around the mean of 3.97 points (*SD*=3.19 points, *Skew*=0.03, *Kurtosis*=-1.49). The average PFT scores in the four conditions did not vary overall, $F(3, 98)=0.510$, $p=.676$, $\eta p^2=.015$. PFT scores were correlated significantly with scores on the Knowledge of Optics post-test ($r[102]=.376$, $p<.001$), so spatial ability was used as a covariate on the analyses for the main outcome measure. PFT scores were not correlated with the Cognitive Load item ($r[101]=-.139$, $p=.165$), so spatial ability was not directly related to how much mental effort participants felt during the computer lesson.

Video Game Experience

Video game experience may also be related to performance on the experimental task and was measured on the prescreening demographics questionnaire. The item on the demographics questionnaire used to approximate video game experience was the participant's self-reported rating for number of hours he/she plays video games each week, because previous research has found that self-reported hours of video game play a week are correlated significantly with both comfort in gaming and measures of video game self-efficacy (Procci, James, & Bowers, 2013). The average number of hours a week participants reportedly played was 4.47 hours (SD=7.28 hours), and hours spent video gaming did not differ by condition, $F(3, 95)=0.220$, $p=.882$, $\eta p^2=.007$. Video game experience as measured by hours of play was not correlated significantly with the main outcome measures, Knowledge of Optics post-test ($r[102]=.173$, $p=.087$) and Cognitive Load ($r[98]=-.064$, $p=.533$).

<u>Brief Assessment of Gesture Survey</u>

Another potential predictor of how well a participant could complete the computer task and subsequent learning was the propensity and perception of gesturing, as measured by the BAG. The BAG consists of four subscales related to a participant's production and perception of gestures: Perception, Production, Social Production, and Social Perception. For each of the subscales, the means and standard deviations are presented in Table 9. The conditions did not differ in any of the four subscales ($F_{Perception}[3, 98]=0.601$, $p=.616$; $F_{Production}[3, 98]=0.902$, $p=.443$; $F_{SocialProduction}[3, 98]=1.326$, $p=.270$; $F_{SocialPerception}[3, 98]=1.546$, $p=.208$). Only the Production subscale was correlated with the main outcome measure, Knowledge of Optics Δ score ($r[102]=0.217$, $p=.029$), which indicated that a participant's propensity for gesture was related to more learning overall. The other three subscales were not correlated with Knowledge of Optics Δ score ($r_{Perception}[102]=0.121$, $p=.225$; $r_{SocialProduction}[102]=0.179$, $p=.072$; $r_{SocialPerception}[102]=0.091$, $p=.361$). Additionally, none of the BAG subscales were correlated significantly with Cognitive Load ($r_{Perception}$[102]=-0.096, p=.342; $r_{Production}$[102]=-0.035, p=..729; $r_{SocialProduction}$[102]=-0.139, p=.165; $r_{SocialPerception}$[102]=-0.088, p=.381).

Table 9. Means and Standard Deviations for the Brief Assessment of Gesture Survey Subscales

Condition		**Perception**		**Production**		**Social Production**		**Social Perception**	
	n	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Total	102	2.10	0.75	3.22	0.75	3.78	0.98	2.57	1.00
Arbitrary Text	23	2.11	0.75	3.45	0.69	4.00	0.75	2.21	1.11
Arbitrary Video	24	2.09	0.57	3.18	0.84	3.46	1.14	2.25	0.81
Natural Text	27	2.25	0.90	3.14	0.78	3.81	1.16	2.56	0.88
Natural Video	28	1.97	0.76	3.15	0.67	3.86	0.77	2.84	1.11

Time Between Prescreening and Experiment

Amount of time between the prescreening and the experiment was measured because there was concern that participants may perform better on the Knowledge of Optics post-test if they were primed on the optics concepts by the pre-test. The average amount of time between participants completing the prescreening online and participating in the lab experiment was 10.16 days (SD=11.23 days), and times were distributed normally ($Skew$=1.49, $Kurtosis$=-0.42). There was a wide range in the amount of time between completing the prescreening and the experiment, from less than a day between prescreening and experiment to 100 days between testing; however, amount of time since the prescreening was not correlated with scores on the Knowledge of Optics post-test (r[102]=-0.02, p=.840), so individuals who completed the pre-test closer in time to the experiment were no more likely to perform better on the post-test. Additionally, the amount of time between prescreening and experiment was not significantly different among conditions, $F(3, 98)=0.773$, p=.512, ηp^2=.023.

Experiment Results

Correlation of Measures

The zero-order correlations between all of the measures are reported in Table 20, Appendix G.

Knowledge of Optics Post-test

Descriptives

The Knowledge of Optics post-test was given after participants completed the computer lesson during the in-lab portion of the experiment. Scores on the post-test were highly correlated with scores on the pre-test ($r[102]=.625$, $p<.001$), such that higher scores on the pre-test were related to higher scores on the post-test. The lowest score was 10.71% correct and the highest score was 71.43%. Like the pre-test scores (M=20.19% correct, SD=12.54%), the Knowledge of Optics post-test scores were distributed normally (*Skew*=0.79, *Kurtosis*=2.11), but the overall average score was higher on the posttest by about 18 percentage points (M=38.72%, SD=18.87%; paired $t[101]=-12.684$, $p<.001$, $d=2.64$). As shown in Table 10, participants scored higher on the post-test than the pre-test with large effect sizes in every condition.

Table 10. Knowledge of Optics Pre-and Post-test, and Time to Complete Computer Lesson Means and Standard Deviations for Each Condition

		Knowledge of Optics (Number Correct)					
		Pre-Test		Post-Test		Pre/Post Difference	
Condition	n	M_A	SD_A	M_B	SD_B	$M_B - M_A$	Cohen's d[a]
Arbitrary Text	23	20.826	12.738	37.273	15.822	16.45	2.58
Arbitrary Video	24	21.268	11.803	37.815	18.904	16.55	2.41
Natural Text	27	20.026	11.716	39.392	20.691	19.37	2.67
Natural Video	28	18.898	14.215	40.026	20.168	21.13	2.75

Note. [a] Cohen's d for repeated measures takes into account the pooled variance of dependent samples

Pre- and Post-test Δ Score

To quantify the extent of participants' conceptual learning between pre-test and post-test on the Knowledge of Optics test and to account for individual differences in pre-test optics knowledge, a Δ score was calculated for each participant by subtracting the pre-test score from the post-test score. The Δ score was then used as the outcome variable for the Knowledge of Optics measure on the following analyses. The average Δ score was a 18.56 percentage points increase between pre- and post-test (SD=14.76, $Skew$=0.36, $Kurtosis$=-1.28). The range of Δ scores was between -10.71, indicating the participant performed 10.71 percentage points worse on the post-test, and a high Δ score of 53.57.

ANCOVAs

A 2X2 between-subjects ANCOVA was conducted on the Knowledge of Optics measure Δ score with continuous PFT score and Video Game Experience as the covariates and two levels of the independent variables: Gesture type (Natural or Arbitrary) and Instruction type (Video or Text). As noted previously, PFT scores and Video Game Experience did not differ by condition. Video Game Experience was not a significant covariate and was removed from the model (F[1, 93]=2.19, p=.143, ηp^2=.023). The covariate, PFT score, was related significantly to the Δ score, F (1, 97)=13.820, p<.001, ηp^2=.125. After accounting for the variance of spatial ability, there were no main effects for either independent variables nor was there an interaction effect (F_G [1, 97]=1.363, p=.246, ηp^2=.014; F_I [1, 97]=0.503, p=.480, ηp^2=.005; F_{G*I}[1, 97]=0.018, p=.892, ηp^2<.001); therefore, only spatial ability was related to learning optics concepts from the computer lesson, and neither type of gesture nor type of instruction affected learning once spatial ability was considered.

Additionally, only one of the four BAG subscales, Production, was correlated significantly with the Knowledge of Optics Δ score, so this variable was used as a covariate in another 2X2 ANCOVA with the same independent variables. PFT was also included as a covariate, which was again a significant predictor of the Δ score (F[1, 96)=11.771, p=.001, ηp^2=.109). BAG Production was marginally significant as a covariate (F[1, 96)=3.915, p=.051, ηp^2=.039). No main effects nor interaction effect were significant (F_I[1, 96]=0.737, p=.393, ηp^2=.008; F_G[1, 96]=2.006, p=.160, ηp^2=.020; F_{G*I}[1, 96]=0.002, p=.969, ηp^2<.001).

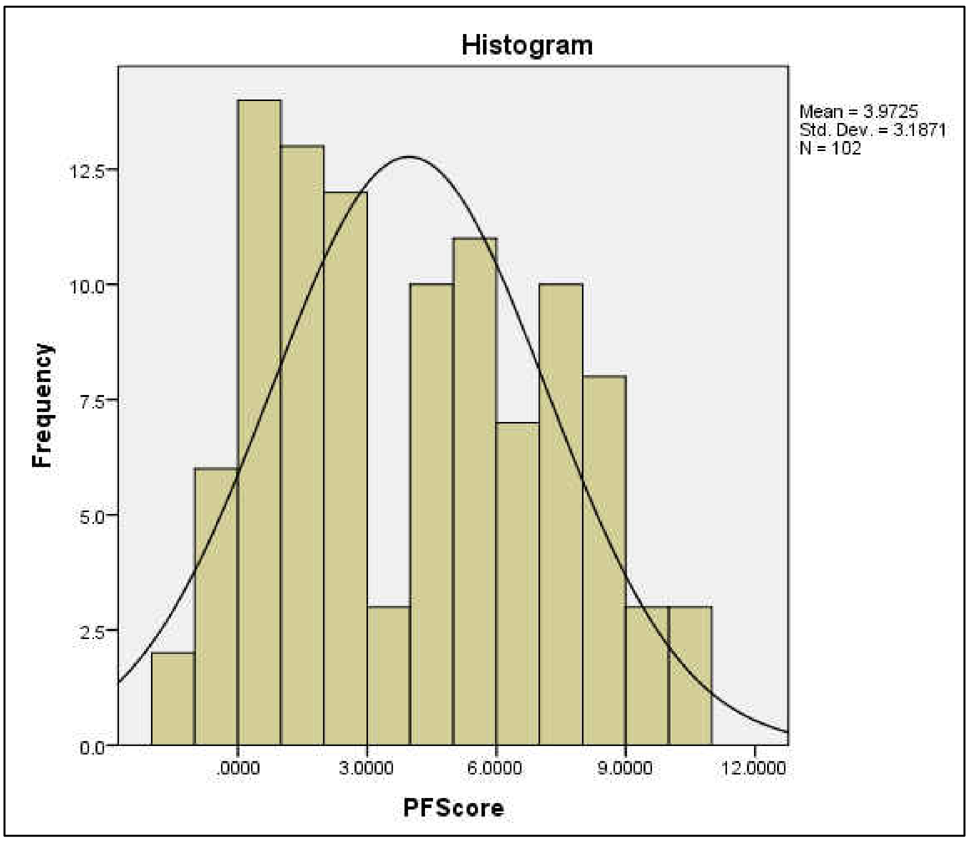


Figure 8. Distribution of PFT Scores. The distribution of scores has a dip in frequency of scores (i.e., number of participants) at the median score of 4.

To better visualize how spatial ability was related to learning from the four conditions, a 4X2 between-subjects ANOVA was conducted with the four conditions (Arbitrary Gesture with Video Instruction, Arbitrary Gesture with Text Instruction, Natural Gesture with Video

Instruction, and Natural Gesture with Text Instruction) and two levels of Spatial Ability (Low and High) on the Δ score. The Low and High Spatial Ability groups were determined by conducting a median split (*Median*=4.0) on the participants' PFT scores. As shown in Figure 8, the distribution of PFT scores had a low frequency (i.e., fewer participants) at the median score of 4 (M=3.97), so there was a distinct division between low and high PFT scores, justifying a median split for Low and High Spatial Ability. Participants who scored 4 or lower on the PFT were considered to have low spatial ability, and those who scored greater than 4 were grouped with higher spatial ability.

As expected based on results from the previous analysis, results of the ANOVA indicated that PFT scores were related to conceptual learning ($F[1, 94]=5.403$, $p=.022$, $\eta p^2=.054$), but Condition itself was not ($F[3, 94]=0.911$, $p=.439$, $\eta p^2=.028$). Whereas in the previous analysis (see ANCOVA above) PFT scores overall did not interact with Condition, when PFT was split into Low and High Spatial Ability, there was a significant interaction between PFT scores and Condition ($F[3, 94]=2.728$, $p=.048$, $\eta p^2=.080$), such that spatial ability determined the degree to which the condition resulted in conceptual learning (Figure 9). Planned comparisons revealed that those with Low Spatial Ability had lower learning gains than those with High Spatial Ability in the Arbitrary Gesture with Text Instruction condition and in the Natural Gesture with Video Gesture condition (Table 11). There were no significant differences between those with Low and High Spatial Ability in either the Arbitrary Gesture with Video Instruction condition or the Natural Gesture with Text Instruction condition.

Table 11. Means, Standard Deviations, and Difference Scores by Condition on the Knowledge of Optics Δ Score

	Knowledge of Optics Delta Score						Difference	
	Low Spatial			High Spatial				
Condition	***n***	***M***	***SD***	***n***	***M***	***SD***	***Cohen's d***	***95% CI***
Arbitrary Text	10	10.00	7.30	13	21.43	12.63	1.07	0.19 – 1.95
Arbitrary Video	17	15.97	16.13	7	17.86	11.48	0.13	-0.76 – 1.01
Natural Text	13	21.15	19.04	14	17.86	14.15	-0.20	-0.95 – 0.56
Natural Video	15	13.33	13.05	13	30.22	13.25	1.29	0.47 – 2.10

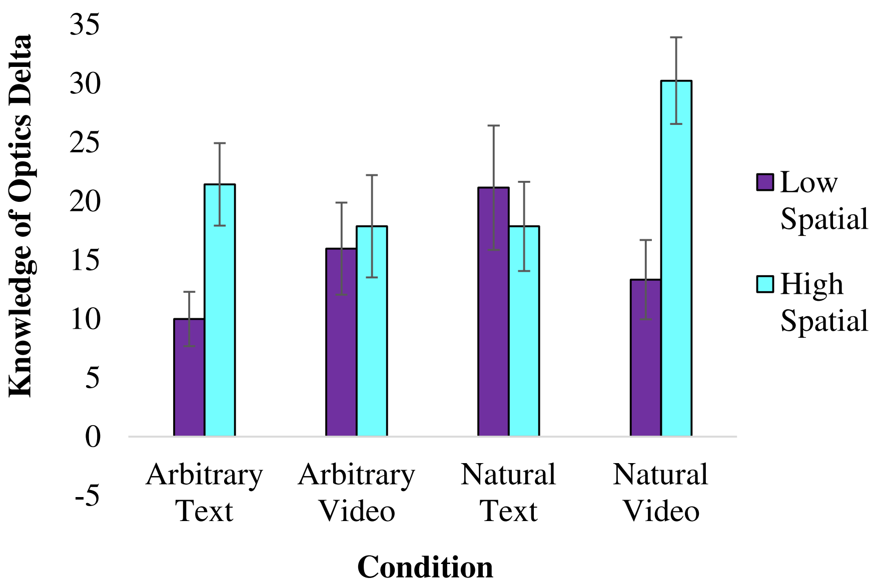


Figure 9. Graph depicting the extent of learning on the Knowledge of Optics Test as measured from pre- to post-test (Δ Score). Those in the Low Spatial Ability group learned significantly less conceptual information than those with in the High Spatial Ability group in the Arbitrary Text condition and the Natural Video condition. Error bars represent standard error.

Cognitive Load

Descriptives

Subjective cognitive load from the computer lesson was assessed by comparing participants' self-reported level of mental effort using the single-item Cognitive Load Questionnaire. Overall, the average mental effort rating was near the middle of the 10-point scale (M=5.46, SD=2.54), with a range from 0-10 and a normal distribution of responses ($Skew$=-0.04, $Kurtosis$=-2.16). The average ratings for each condition are presented in Table 12.

Table 12. Cognitive Load Rating by Condition

		Mental Effort Ratings	
Condition	***n***	***M***	***SD***
Arbitrary Text	22[a]	7.14	1.98
Arbitrary Video	24	6.54	1.79
Natural Text	27	4.26	2.81
Natural Video	28	4.36	2.11

Note. [a] One participant did not respond to this item

ANCOVA

To test whether cognitive load differed depending on type of gesture interaction or instruction, another 2X2 between-subjects ANCOVA was conducted on the mental effort rating with spatial ability and video game experience as the covariates. Spatial ability, as measured by the PFT, and video game experience were included in the analysis to mirror the analysis of the Knowledge of Optics measure and provide a complete picture of the variables of interest; although, as previously discussed, PFT and Video Game Experience were not correlated with the cognitive load measure nor did they differ overall by condition.

Video Game Experience was not a significant covariate (F[1, 92]=0.863, p=.355, ηp^2=.009) and accounted for very little variance, so it was removed from the analysis. The results of the ANCOVA for the cognitive load measure differed from those of the previous analysis for Knowledge of Optics in that spatial ability was not a significant covariate for cognitive load (F[1, 96]=2.114, p=.149, ηp^2 =.022). As shown in Figure 10, there was a significant main effect for Gesture (F[1, 96]=31.859, p<.001, ηp^2=.249), such that those using natural gestures to interact with the computer lesson felt less mental effort (M=4.31, SD=2.46) than those using arbitrary gestures (M=6.83, SD=1.89) by about 25%. There was not a significant main effect for Instruction (F[1, 96]=0.531, p=.468, ηp^2=.006) nor was there an interaction effect (F[1, 96]=0.748, p=.389, ηp^2=.008).

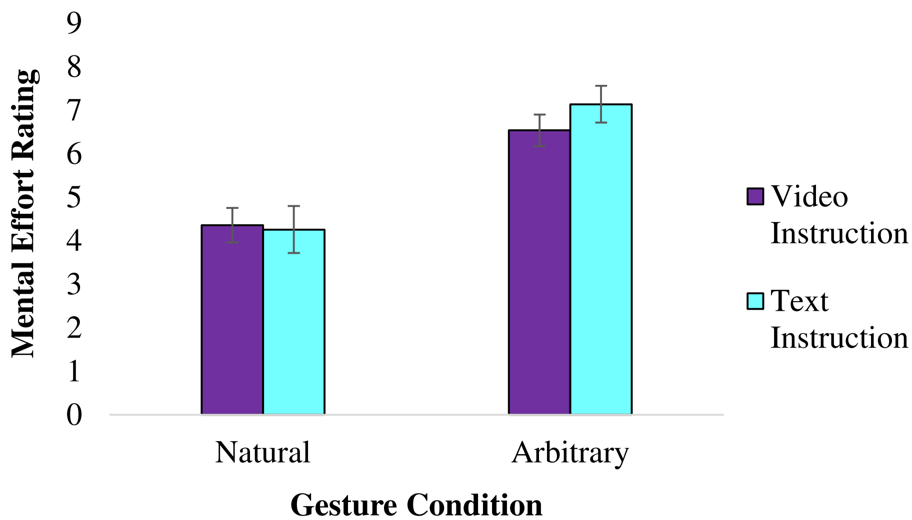


Figure 10. Graph depicting perceived cognitive load as measured by the ratings on the mental effort rating scale, with a higher rating indicating more mental effort to complete the computer lesson. Participants in the Natural Gesture conditions felt less cognitive load than those in the Arbitrary Gesture conditions. Error bars represent standard error.

Instructional Efficiency

Instructional efficiency is an approach for comparing different instruction types that considers the learning gain in conjunction with the amount of mental effort expended during the lesson (Sweller et al., 1998). The formula for instructional efficiency (E) creates a relative measure of cognitive load and performance by converting both mental effort rating and performance measure into z scores (Paas & van Merriënboer, 1994):

$$E = \frac{Z_{\text{Mental Effort}} - Z_{\text{Performance}}}{\sqrt{2}}$$

when $Z_{\text{Mental Effort}} - Z_{\text{Performance}} < 0$, then E is positive
when $Z_{\text{Mental Effort}} - Z_{\text{Performance}} > 0$, then E is negative

To determine whether the conditions differed in instructional efficiency, an instructional efficiency value was calculated for participants using the mental effort rating and the Knowledge of Optics Δ score. The means and standard deviations for the instructional efficiency values are reported in Table 13.

Table 13. Means and Standard Deviations of Instructional Efficiency Scores

		Instructional Efficiency Scores (*E*)	
Condition	***n***	***M***	***SD***
Arbitrary Text	23	-0.57	0.83
Arbitrary Video	24	-0.40	0.96
Natural Text	27	0.38	1.16
Natural Video	28	0.43	0.94

A 2X2 between-subjects ANCOVA was then conducted on the instructional efficiency scores *E* using the two independent variables (Gesture type and Instruction type). In testing the assumptions of ANCOVA, Levene's Test of Equality of Error Variance was significant (F[3, 97]=4.929, p=.003); however, Tabachnick and Fidell (2013) propose using the F_{Max} ratio to determine whether unequal sample variances violate the ANCOVA assumption of homogeneity of variance to the extent that other analyses or data transformations should be performed (p. 86). F_{Max} is a ratio of the largest sample variance to the smallest sample variance. For relatively equal sample sizes (i.e., less than 4 to 1 ratio of cell sample sizes), they suggest that a F value up to 10 is acceptable. Because the ratio between the sample variance of the largest variance and that of the smallest variance did not exceed the F_{Max} value of 10 ($\sigma^2_{NaturalVideo}$=0.55, $\sigma^2_{NaturalText}$=1.33; F_{Max}=2.43), no adjustments were made.

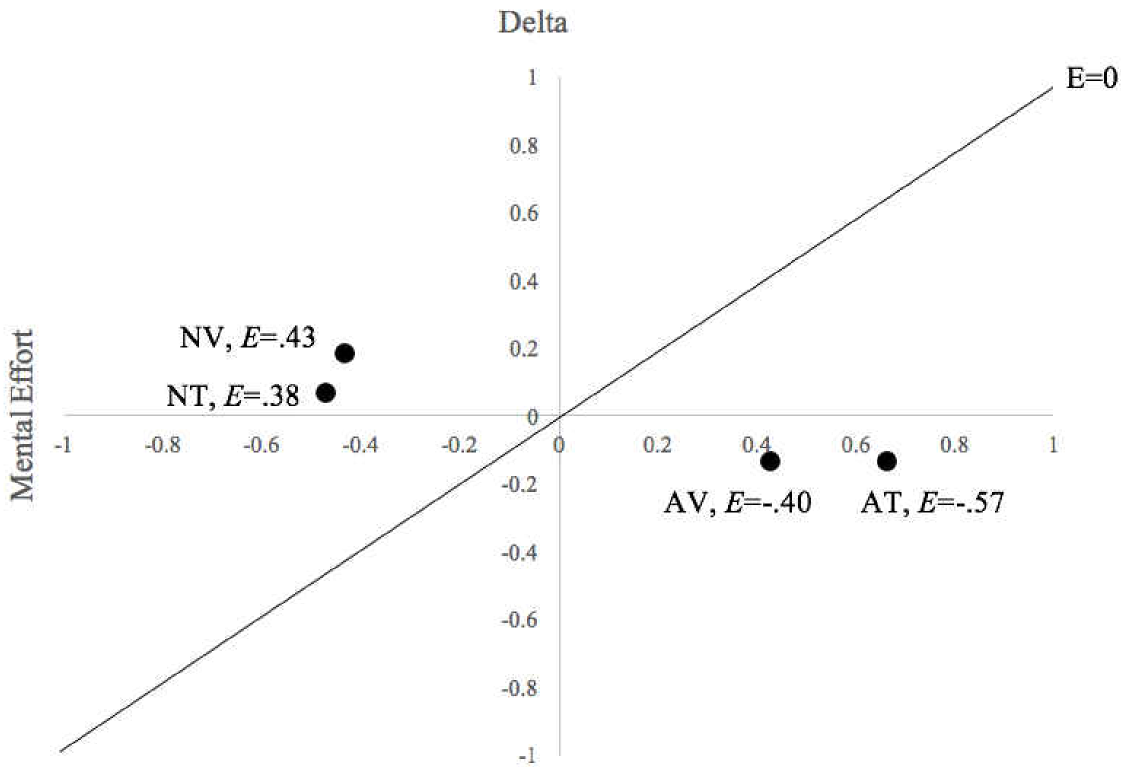


Figure 11. Graph depicting the instructional efficiency (*E*) of the conditions. Conditions plotted above the line (*E*=0) are more efficient, and conditions below the line are less efficient. AT = Arbitrary Gesture with Text Instruction condition; AV = Arbitrary Gesture with Video Instruction condition; NT = Natural Gesture with Text Instruction condition; NV = Natural Gesture with Video Instruction condition.

Results of the ANCOVA indicated that PFT was a significant covariate ($F[1, 96]=14.742$, $p<.001$, $\eta p^2=.133$), so spatial ability was related to instructional efficiency. There was also a main effect for Gesture type ($F[1, 96]=23.625$, $p<.001$, $\eta p^2=.197$), such that Natural Gestures ($M=0.44$, $SD=0.96$) were significantly more efficient than Arbitrary Gestures ($M=-0.47$, $SD=0.89$). There was not a main effect for Instruction ($F[1, 96]=1.109$, $p=.295$, $\eta p^2=.011$), nor was there an interaction effect ($F[1, 96]=0.259$, $p=.612$, $\eta p^2=.003$). The results of this ANCOVA suggest that while spatial ability is a significant predictor of instructional efficiency, the extent of instructional efficiency is determined mostly by the type of interaction with the computer system. Figure 11 depicts this relationship, with conditions that are instructionally efficient plotted above the diagonal line, and conditions that are less efficient plotted below the line.

ANOVA

Because the analyses indicated spatial ability was a predictor of instructional efficiency, a follow-up analysis was conducted to understand how instructional efficiency differs in the conditions depending on participant's spatial ability in a 2 Gesture (Natural or Arbitrary) X 2 Instruction (Video or Text) X 2 Spatial Ability (High or Low) between-subjects ANOVA. Levene's Test was significant ($F[7, 93]=3.038$, $p=.006$); however, the F_{Max} ratio again did not exceed 10, and no adjustments were made ($\sigma^2_{ArbitraryText}=0.217$, $\sigma^2_{NaturalTextt}=1.796$; $F_{Max}=8.276$). There was a main effect for PFT ($F[1, 93]=9.431$, $p=.003$, $\eta p^2=.092$), in which all of the conditions were more instructionally efficient for those with High Spatial Ability ($M=.296$, $SD=.856$) than Low Spatial Ability ($M=-.248$, $SD=1.100$). Additionally, there was a main effect for Gesture type ($F[1, 93]=22.075$, $p=<.001$, $\eta p^2=.192$), where Natural Gestures ($M=.404$, $SD=.959$) were again more instructionally efficient than Arbitrary Gestures ($M=-.471$, $SD=.900$). There was not a main effect for Instruction ($F[1, 93]=1.428$, $p=.235$, $\eta p^2=.015$), nor were there any interaction effects ($F_{I*G}[1, 93]=0.613$, $p=.436$, $\eta p^2=.007$; $F_{I*PFT}[1, 93]=0.000$, $p=.989$, $\eta p^2<.001$; $F_{G*PFT}[1, 93]=1.909$, $p=.170$, $\eta p^2=.020$; $F_{I*G*PFT}[1, 93]=1.696$, $p=.196$, $\eta p^2=.018$). These results are graphed in Figure 12.

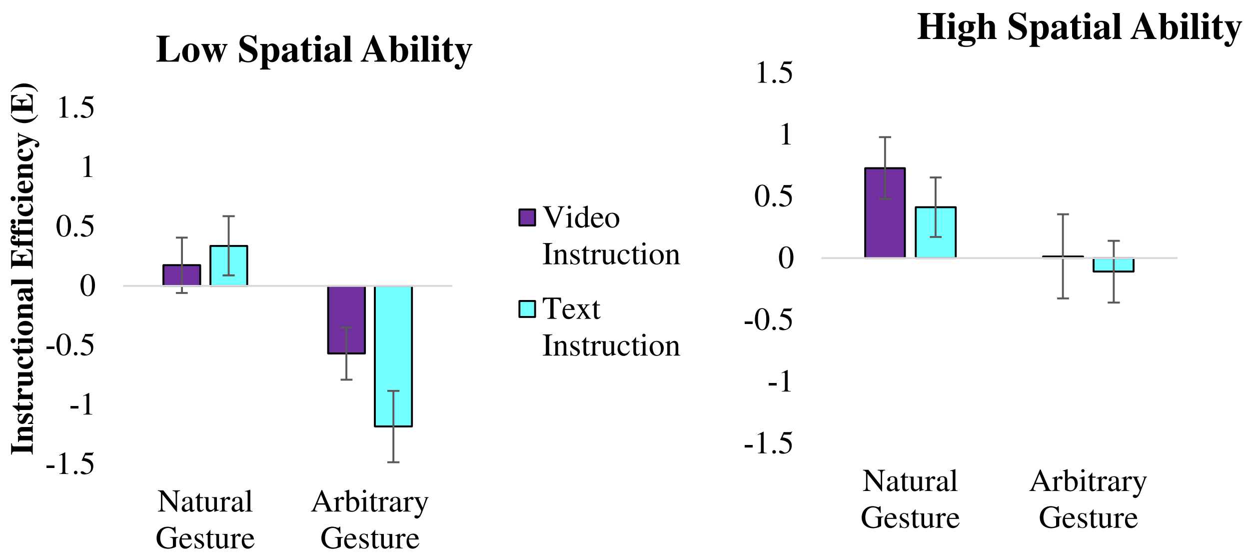


Figure 12. Graph depicting the instructional efficiency (*E*) of conditions by spatial ability. When *E* is above 0, the instruction was efficient. There was a main effect for Spatial Ability and Gesture type. There were no other main effects or any interaction effects. Error bars represent standard error.

<u>Time in Computer Lesson</u>

Descriptives

The total time to complete the computer lesson was recorded for each participant in minutes. The minimum time in which a participant completed the lesson was 8.09 minutes, and the maximum time to complete the lesson was 15.24 minutes. Overall, the average time participants spent in the experimental testbed was 10.64 minutes (SD=1.69). The distribution of times was skewed right ($Skew$=4.74, $Kurtosis$=1.12). Table 14 shows the average time each condition took to complete the computer lesson.

Table 14. Time to Complete Computer Lesson by Condition

		Time (minutes)	
Condition	***n***	***M***	***SD***
Arbitrary Text	23	12.25	1.80
Arbitrary Video	24	10.89	1.43
Natural Text	27	9.66	0.87
Natural Video	28	10.10	1.44

ANCOVA

A 2X2 between-subjects ANCOVA was conducted on amount of time to complete the computer lesson with PFT as the covariate and two levels of the independent variables: Gesture type (Natural or Arbitrary) and Instruction type (Video or Text). Levene's Test of Equality of Error Variance was significant (F[3, 98]=4.145, p=.008), but the F_{Max} ratio was again below the accepted value of 10 ($\sigma^2_{NaturalText}$=0.76, $\sigma^2_{ArbitraryText}$=3.24; F_{Max}=4.26), and no adjustments were made. The unstandardized residuals for the ANCOVA were skewed right (*Shapiro-Wilk*=.932, p<.001). To reduce the positive skew, I performed a log transformation of the time variable as suggested by Field (2013) and re-conducted the analysis. The results of the ANCOVA on the transformed data were nearly identical to the results of the un-transformed data: The previously non-significant variables remained non-significant and the same was true of significant variables, with miniscule increases to the ηp^2 values. Because the transformation resulted in inconsequential changes to the ANCOVA model, and the transformed variable is less interpretable than raw scores (Field, 2013), I decided to report the original ANCOVA results below with raw time scores (i.e., minutes to complete the computer lesson).

The ANCOVA results indicated that PFT was not a significant covariate (F[1, 97]=1.733, p=.191, ηp^2=.018). There was an interaction effect for Gesture and Instruction (F[1, 97]=10.765, p=.001, ηp^2=.100). Although Field (2013) suggests it may not be appropriate to address a main effect when an interaction effect exists, the main effect for Gesture accounted for a large amount of the variance in the time taken to complete the computer lesson (F[1, 97]=35.854, p<.001, ηp^2=.270], with the Natural Gesture group (M=9.89, SD=1.20) finishing the computer lesson almost two minutes faster than the Arbitrary Gesture group (M=11.56, SD=1.75). Simple effects contrasts (see Table 14 for descriptives) showed that there was also a difference in the Arbitrary Gesture group with those getting Video Instruction finishing the computer lesson faster than the Text Instruction group by approximately one minute (d=0.839). The difference in time to complete the lesson for those with Natural Gestures was not significant between Video and Text Instruction groups (d=0.368).

System Usability Scale

Descriptives

Participant ratings on the SUS items were averaged to determine the perceived usability of the computer lesson in each condition. The scores on the SUS (α=.86) were distributed normally (*Skew*=-2.08, *Kurtosis*=-1.31) around the mean of 3.728 (SD=0.761). The range of average SUS scores was between 1.80 and 5.00 on the 7-point Likert-type scale, with higher scores indicating more perceived usability. Means and standard deviations for each condition are presented in Table 15.

Table 15. System Usability Scale Table of Means and Standard Deviations for Each Condition

	SUS Ratings		
Condition	***n***	***M***	***SD***
Arbitrary Text	23	3.257	0.862
Arbitrary Video	24	3.354	0.678
Natural Text	27	4.159	0.491
Natural Video	28	4.018	0.592

ANCOVA

To determine whether SUS ratings differed by condition, a 2X2 between-subjects ANCOVA was conducted with SUS scores as the DV, PFT as the covariate, and two levels of each independent variable: Gesture type (Natural or Arbitrary) and Instruction type (Video or Text). Levene's Test of Equality of Error Variance was significant ($F[3, 98]=3.624$, $p=.016$); however, the F_{Max} ratio did not exceed 10, so no adjustments were made ($\sigma^2_{NaturalText}=0.241$, $\sigma^2_{ArbitraryText}=0.743$; $F_{Max}=3.08$). PFT was not a significant covariate ($F[1, 97]=0.023$, $p=.881$, $\eta p^2<.001$). There was a main effect for type of Gesture ($F[1, 97]=35.148$, $p<.001$, $\eta p^2=.266$), where those interacting with Natural gestures ($M=4.087$, $SD=0.544$) rated the system higher in usability than those using Arbitrary gestures ($M=3.306$, $SD=0.767$). There was not a main effect for Instruction ($F[1, 97]=0.022$, $p=.881$, $\eta p^2<.001$), nor an interaction between Gesture and Instruction ($F[1, 97]=0.834$, $p=.363$, $\eta p^2=.009$).

Presence Questionnaire

Descriptives

The PQ consists of four subscales measuring different dimensions of sense of presence felt by participants during the computer lesson: Involvement (α=.84), Sensory Fidelity (α=.67), Adaptation/Immersion (α=.70), and Interface Quality (α=.70). As described in the Materials section in Chapter 6, the PQ was used because each of these factors may be related to how much a participant learned from the computer lesson. Provided in Table 16 are the means and standard deviations for each subscale on a 7-point scale, with higher scores on the first three subscales indicating more presence as related to each dimension (i.e., higher feeling of involvement, sensory fidelity, or immersion, respectively). Interface Quality differs from the other subscales in that a higher score indicate less presence because it measures the extent to which the computer system distracts from performance in the virtual environment – that is, a higher score indicates lower interface quality.

Table 16. Means and Standard Deviations for the Presence Questionnaire Subscales

Condition		**Involvement**		**Sensory Fidelity**		**Adaptation/ Immersion**		**Interface Quality**[a]	
	n	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Total	102	4.68	0.99	5.43	1.13	4.96	0.93	3.07	1.18
Arbitrary Text	23	4.55	1.01	5.43	0.91	4.66	0.91	3.39	1.44
Arbitrary Video	24	4.23	0.94	5.17	1.34	4.65	1.06	3.39	0.94
Natural Text	27	5.00	0.87	5.43	1.11	5.37	0.85	2.99	1.27
Natural Video	28	4.87	1.02	5.64	1.14	5.07	0.75	2.62	0.87

Note. [a] Higher scores indicate *worse* interface quality. The subscale of Interface Quality differs from the other subscales in that a higher score reflects less presence.

ANOVAs

First, an ANOVA was conducted for each subscale to determine whether that dimension of presence differed as a function of the conditions. In each 2X2 ANOVA, the PQ subscale was the outcome variable, and there were two types of Gesture interaction (Natural or Arbitrary) and two types of Instruction (Video or Text). All of the PQ subscales had normally distributed residuals (*Shapiro-Wilk* ps>.05).

For the Involvement subscale, there was a main effect for Gesture type ($F[1, 98]=8.124$, $p=.005$, $\eta p^2=.077$), indicating that participants in the Natural Gesture conditions rated the control of the computer interface as more natural than the Arbitrary Gesture conditions. Instruction type ($F[1, 98]=1.378$, $p=.243$, $\eta p^2=.014$) and the interaction of Gesture and Instruction ($F[1, 98]=0.252$, $p=.617$, $\eta p^2=.003$) were not significant predictors of the Involvement subscale.

The Sensory Fidelity subscale, which measures the extent to which the senses are engaged with the computer system, was not predicted by either Gesture type ($F[1, 98]=1.106$, $p=.295$, $\eta p^2=.011$), Instruction type ($F[1, 98]=0.016$, $p=.900$, $\eta p^2<.001$), or their interaction ($F[1, 98]=1.126$, $p=.291$, $\eta p^2=.011$).

In the Adaptation/Immersion subscale that measures the participants' perceived ability to concentrate on or be immersed in the computer task, there was a main effect for Gesture ($F[1, 98]=10.223$, $p=.002$, $\eta p^2=.094$). Those in the Natural Gesture conditions rated their sense of immersion higher than those in the Arbitrary Gesture conditions, although all conditions were above the midpoint (i.e., 3) on the scale, indicating high immersion overall. There was not a main effect for Instruction type ($F[1, 98]=0.767$, $p=.383$, $\eta p^2=.008$), nor was there an interaction effect ($F[1, 98]=0.663$, $p=.417$, $\eta p^2=.007$).

Finally, the Interface Quality subscale was compared by condition. This was the only subscale in which a lower score indicated higher presence in terms of better interface quality. Levene's Test was significant ($F[3, 98]=3.80$, $p=.013$), but the F_{Max} ratio was below the acceptable value of 10 ($\sigma^2_{NaturalVideo}=0.755$, $\sigma^2_{ArbitraryText}=2.074$; $F_{Max}=2.75$), and no adjustments were made. Results of the ANOVA indicated there was a main effect for Gesture type (F[1, 98]=6.672, p=.011, ηp^2=.064), such that those in the Natural Gesture conditions felt the interface quality was better quality than those in the Arbitrary Gesture conditions. Neither Instruction type (F[1, 98]=0.666, p=.417, ηp^2=.007) nor the interaction between Gesture and Instruction (F[1, 98]=0.647, p=.423, ηp^2=.007) were significant predictors of Interface Quality.

ANCOVA

The Sensory Fidelity subscale was the only PQ subscale that did not differ by condition, so it was included in an additional analysis as a covariate of the Knowledge of Optics outcome variable (Δ Score), to determine whether sensory fidelity affected how much participants learned in each condition. A 2X2 ANCOVA was conducted on the two Gesture conditions (Natural or Arbitrary) and the two Instruction conditions (Video or Text), with PFT and Sensory Fidelity as covariates. Although PFT was a significant covariate (F[1, 96]=13.843, p<.001, ηp^2=.126), Sensory Fidelity was not a significant covariate when spatial ability was included (F[1, 96]=1.66, p=.201, ηp^2=.017). Just as reported in previous analyses on the Δ Score, no other main effects or interaction effects were significant (F_I[1, 96]=0.528, p=.469, ηp^2=.005; F_G[1, 96]=1.061, p=.306, ηp^2=.011; F_{I*G}[1, 96]<0.001, p=.999, ηp^2<.001).

Summary of Results

The hypotheses tested in the results sections were mostly not supported, with the exception of Natural Gesture conditions having lower perceived cognitive load than the Arbitrary Gesture conditions. A summary of whether each hypothesis was supported or not is presented in Table 17.

Table 17. Summary of Support for Hypotheses

	Hypotheses	Supported (Yes/No)
H1.1	Main effect for Gesture type on Knowledge of Optics Δ *Natural Gesture > Arbitrary Gesture*	No
H1.2	Covariate for Spatial Ability on Knowledge of Optics Δ	Yes
H1.3	Main effect for Gesture type on Cognitive Load *Natural Gesture < Arbitrary Gesture*	Yes
H1.4	Covariate for Spatial Ability on Cognitive Load	No
H1.5	Main effect for Gesture type on Instructional Efficiency *Natural Gesture > Arbitrary Gesture*	Yes
H2.1	Main effect for Instruction type on Knowledge of Optics Δ *Video Instruction > Text Instruction*	No
H2.2	Main effect for Instruction type on Cognitive Load *Video Instruction < Text Instruction*	No
H2.3	Main effect for Instruction type on Instructional Efficiency *Video Instruction > Text Instruction*	No
H3.1	Simple effect for Instruction and Gesture on Knowledge of Optics Δ *Natural Gesture with Video Instruction > Other conditions*	No
H3.2	Simple effect for Instruction and Gesture on Cognitive Load *Natural Gesture with Video Instruction < Other conditions*	No
H3.3	Simple effect for Instruction and Gesture on Knowledge of Optics Δ *Arbitrary Gesture with Text Instruction < Other conditions*	No
H3.4	Simple effect for Instruction and Gesture on Cognitive Load *Arbitrary Gesture with Text Instruction > Other conditions*	No
H3.5	Main effect for Spatial Ability on Instructional Efficiency *High Spatial > Low Spatial*	Yes

Additionally, the results were clarified by further analyses that included the potential covariates (e.g., spatial ability), and these analyses are summarized in Table 18. The summary of the additional research questions are divided into those that relate to the prescreening analyses and experimental analyses. The prescreening analyses are presented first, summarizing whether the prescreening variables differed by condition. These prescreening variables were analyzed to determine if they differed by condition because they were likely predictors of the experimental outcome variables and inequalities among the conditions could bias results. The lower section of the table summarizes whether the subject variables predicted the experimental outcome measures and other analyses that provide insight into how the conditions differed.

Table 18. Summary of Results from Additional Research Questions

Research Questions		Analysis	Results	Answer (Yes/No)
Prescreening				
1.	Does conceptual knowledge about optics (i.e., Knowledge of Optics Pre-test) differ by condition before the experiment?	ANOVA	Main effect for Condition not significant	No
2.	Does spatial ability (i.e., PFT) differ by condition?	ANOVA	Main effect for Condition not significant	No
3.	Does video game experience (i.e., hours per week playing video games) differ by condition?	ANOVA	Main effect for Condition not significant	No

4.	Does propensity to and perception of gesturing differ by condition (i.e., BAG subscales)?	ANOVA for each subscale	A) *BAG Perception* Main effect for Condition not significant B) *BAG Production* Main effect for Condition not significant C) *BAG Social Production* Main effect for Condition not significant D) *BAG Social Perception* Main effect for Condition not significant	No (for all subscales)
5.	Does the time between the online prescreening and in-lab experiment differ by condition?	ANOVA	Main effect for Condition not significant	No
6.	Does the time between the online prescreening and the in-lab experiment relate to conceptual learning (i.e., Knowledge of Optics Δ score)?	Correlation	Time between Prescreening- Experiment not correlated with Knowledge of Optics Δ score	No
Experiment				
7.	Does spatial ability (i.e., PFT) predict conceptual learning (i.e., Knowledge of Optics Δ score)?	ANCOVA	PFT significant covariate	Yes
8.	Does video game experience (i.e., hours per week playing video games) predict conceptual learning (i.e., Knowledge of Optics Δ score)?	ANCOVA	Video game hours per week not a significant covariate	No

9.	Does conceptual learning (i.e., Knowledge of Optics Δ score) differ as a function of spatial ability (High or Low) and condition?	ANOVA	PFT X Condition interaction effect significant (High Spatial Ability had higher Δ score than Low Spatial Ability in Arbitrary Text & Natural Video conditions)	Yes (2 of 4 conditions)
10.	Does spatial ability (i.e., PFT) predict cognitive load (i.e., mental effort rating)?	ANCOVA	PFT not a significant covariate	No
11.	Does video game experience (i.e., hours per week playing video games) predict cognitive load (i.e., mental effort rating)?	ANCOVA	Video game hours per week not a significant covariate	No
12.	Does propensity to and perception of gesturing (i.e., BAG Production) predict conceptual learning (i.e., Knowledge of Optics Δ score)?	ANCOVA	BAG Production marginally significant covariate (p=.051)	Yes
13.	Does instructional efficiency differ by condition?	ANCOVA	Main effect for Gesture significant (Natural > Arbitrary)	Yes
14.	Does time to complete the computer lesson (in minutes) differ by condition?	ANOVA	Gesture X Instruction interaction effect significant (Arbitrary Video faster than Arbitrary Text condition, & both Natural Gesture faster than Arbitrary Gesture)	Yes
15.	Does spatial ability (i.e., PFT) predict time to complete the computer lesson (in minutes)?	ANCOVA	PFT not a significant covariate	No
16.	Does perceived usability of the computer lesson (i.e., SUS) differ by condition?	ANOVA	Main effect for Gesture significant (Natural > Arbitrary)	Yes

17.	Does spatial ability (i.e., PFT) predict perceived usability of the computer lesson (i.e., SUS)?	ANCOVA	PFT not a significant covariate	No
18.	Does sense of presence in the computer lesson (i.e., PQ subscales) differ by condition?	ANOVA for each subscale	A) *PQ Involvement* Main effect for Gesture significant (Natural > Arbitrary) B) *PQ Adaptation/ Immersion* Main effect for Gesture significant (Natural > Arbitrary) C) *PQ Sensory Fidelity* Main effects for Gesture & Instruction not significant Interaction effect not significant D) *PQ Interface Quality* Main effect for Gesture significant (Natural < Arbitrary[a])	Yes (3 of 4 subscales)
19.	Does sense of presence in the computer lesson (i.e., PQ Sensory Fidelity) predict conceptual learning (i.e., Knowledge of Optics Δ score)?	ANCOVA	PQ Sensory Fidelity not a significant covariate	No

Note. [a] PQ Interface Quality is scored such that lower scores indicate better interface quality and presence

Participant Reactions

Participants responded to an open-ended question following the SUS asking them to write any comments related to the computer lesson. In general, participants wrote positively about their experience using the computer system, even in conditions associated with more mental effort, lower usability ratings, and longer time to complete the lesson. For example, Table 19 lists representative responses for each condition, and these responses reflect the finding that usability was high overall (i.e., every participant rated the system above the mid-point on the usability scale).

Table 19. Participant Perceptions of the Computer System

Condition	*Perception of Computer System*
Arbitrary Gesture with Text Instruction	"I really enjoyed using gesture commands to walk through this experiment. Would 100% use again."
Arbitrary Gesture with Video Instruction	"Enjoyed this study!"
Natural Gesture with Text Instruction	“Very fun to do, feels very futuristic"
Natural Gesture with Video Instruction	“You did an amazing job designing the system. Well done!"

In addition to these general comments, other participant responses reflected the empirical results reported above. Specifically, some participants in the arbitrary conditions reported not being able to focus on the learning material (i.e., optics concepts) because they were focused on remembering the gestures, supporting the results of the experiment in which those using arbitrary gestures felt more cognitive load than those using natural gestures to interact:

> "The system was easy to use once I adjusted to the required gestures. The difficulty was remembering them all, particularly the ones that were not simply inverted (eg clockwise and counter clockwise). As a result of this learning curve I wasn't always able to focus on the information as I was too focused on learning the interface. With that said, once the interface is learned it all began to go smoothly [sic]." – from Arbitrary Gestures with Video Instruction condition

"I felt completly focused on getting the gestures correct that I wasn't really focused on the material I was supposed to be learning [sic]." – from Arbitrary Gestures with Video Instruction condition

Other comments reflect the results of the amount of time it took to complete the computer lesson, mirroring the empirical results showing those in the arbitrary gesture conditions took longer to complete the lesson. For example, a participant in the Arbitrary Gestures with Text Instruction condition commented, "The system is workable, just takes abit of time [sic]." Finally, other participants proposed changes to the system in their comments that were related to the research questions of the study – that is, are gestures seen as more natural if they relate to the instructional material, resulting in higher learning outcomes, and does how the gestures are instructed help understanding of the gestures.

"Make the gestures and the actions more relatable." – from Arbitrary Gestures with Text Instruction condition

"A picture of the gestures or demonstration would be helpful prior to the task." – from Arbitrary Gestures with Text Instruction condition

"I think there needs to be more percision in the motion tracker so that smaller gestures are still registered [sic]" – from Natural Gestures with Video Instruction condition

CHAPTER EIGHT: DISCUSSION

Overview

As gesture-based interactions with computer interfaces become more technologically feasible for educational and training systems, it is important to consider what interactions are best for the learner. Computer interactions should not interfere with learning nor increase the mental effort of completing the lesson. The purpose of the current set of studies was to determine whether type of gesture-based interaction, or instruction of those gestures, affects the learner in a computer lesson. To test whether the type of interaction affects conceptual learning in a computer lesson, the gesture-based computer interactions were either naturally- or arbitrarily-mapped to the learning material. The natural gestures implemented in the computer lesson were those that were performed in Study 1 and rated in Study 2 as most closely resembling the physical interaction they represent. The arbitrary gestures were also rated by participants as most arbitrary for each computer action in Study 2. To test whether the effect of novel gesture-based interactions depends on how they are taught, the way the gestures were instructed was varied in the main experiment by using either video- or text-based tutorials.

Based on the theoretical frameworks of Embodied Cognition and CLT, it was hypothesized that using natural gestures to interact with the computer lesson would increase how much conceptual information was learned while decreasing the amount of mental effort felt during the lesson, and arbitrary gestures would have the opposite effects. This is because natural gestures help develop sensorimotor mental representations for the conceptual information as well as reduce the extraneous load of interacting with the computer lesson. In contrast, arbitrary gestures that do not match the learning material do not help the development of a schema for the conceptual information because the gestures do not have information about the lesson that may help understanding of the concept and would serve to add extraneous information that must be processed in working memory unrelated to the lesson.

Furthermore, it was also predicted that instructing the gestures using video-based tutorials would lead to more conceptual learning and less mental effort than text-based tutorials. This prediction was based on the Embodied Cognition paradigm that would suggest video tutorials for learning interactions may help learners understand the new gestures by providing visualization of the gestures, which activates the sensorimotor system such that gestures can later be recalled in the same sensorimotor state. Conversely, text-based tutorials may not activate the sensorimotor system if they are processed as verbal information, and the learner would not have the encoding and recall benefit of priming the sensorimotor system. Additionally, it was hypothesized that there would be a combined effect of gesture and instruction type, such that natural gesturing with video instructions would be the best condition for learning and mental effort, while arbitrary gestures with text instructions would be the worst condition. Finally, it was predicted that individual differences of the participants, most notably spatial ability, would impact the amount learned from the computer lesson.

I tested each of these hypotheses using a crossed experimental design in which participants were assigned to one of four conditions. Results of the experiment, which are discussed in depth in this chapter, support the overarching theme that natural gesture-based interactions were better for learning than arbitrary gestures; however, instruction of the gestures did not affect learning or how much mental effort was felt during the task. Furthermore, there was not an interaction of the manipulated conditions. Results were also largely dependent on an individual's spatial ability, such that the instructional efficiency of the conditions differed by high and low spatial ability. These findings and their implications for instructional design theory and practice are discussed below.

Conceptual Learning

Before considering the instructional efficiency of each factor, which is a relative measure of learning that takes into consideration the mental effort involved, I first analyzed the extent of conceptual learning that occurred. Conceptual learning was measured using a Knowledge of Optics test that was developed to quantify prior knowledge of optics concepts (pre-test) and how much knowledge was gained from the computer lesson (post-test). The learning gain was calculated for each participant by subtracting the pre-test score from the post-test score to create a knowledge of optics learning score (Δ). Overall, every condition showed a large learning gain between the pre-test and the post-test after the computer lesson, with an average gain of 18 percentage points.

Next, to test whether type of gesture, type of instruction, or their interaction were predictive of the learning gain, an ANCOVA was conducted. Spatial ability was also included as a potential predictor in the analysis. The results indicated that participants' spatial ability accounted for most of the differences in learning gain. Because spatial ability predicted most of the learning gain, neither gesture nor instruction types impacted conceptual learning.

To elucidate why spatial ability was the most significant predictor of learning, a follow-up analysis was conducted to see if amount of spatial ability interacted with the conditions to explain conceptual learning gains. Participants were divided into high and low spatial ability groups at the median score of the Paper Folding Test (this distribution was bimodal at the median). When amount of spatial ability was analyzed in conjunction with the conditions, there was an interaction between spatial ability and the conditions. This interaction indicated that in two of the four conditions, Arbitrary Gesture with Text Instruction condition and Natural Gesture with Video Gesture condition, participants with low spatial ability had lower learning gains than those with high spatial ability. In two of the conditions, Arbitrary Gesture with Video Instruction condition and Natural Gesture with Text Instruction condition, there was not a difference between those with low and high spatial ability on conceptual learning. These findings partially support the hypotheses that 1. Natural Gestures with Video Instruction condition would learn the most, and 2. Arbitrary Gestures with Text Instructions would learn the least, but the first hypothesis was only true for those with high spatial ability and the second hypothesis was only true for those with low spatial ability.

The difference between those with low and high spatial ability in two of the four conditions may be explained in two ways. First, the Natural Video condition did have the highest conceptual learning, but only for those with high spatial ability. Participants with low spatial ability performed on par with those with low spatial ability in the other conditions. This suggests that spatial ability may enhance learning when natural gestures and video instructions are combined, lending evidence to the spatial ability-as-enhancer hypothesis (Mayer & Sims, 1994). This hypothesis states that when the instruction is good (i.e., the combination of natural gestures and video instructions), those with higher spatial ability will benefit while those with lower spatial ability must spend more cognitive effort either creating schemas or making representational connections. Extending from the spatial ability-as-enhancer hypothesis, the added benefit that participants with high spatial ability receive from good instruction may be related to their ability to create mental animations related to schema development of the conceptual material. There are two parts to this argument: 1. Identifying good instruction, and 2. Explaining why those with high spatial ability may be better at mental animation.

The instruction in this condition might be considered good because the natural gestures are directly related to the learning material, and therefore, may not pose the additional burden on working memory of remembering information that is not task-relevant. Natural gestures may also serve as a cue for later recall of the conceptual information related to those actions – that is, mental animations in the conceptual schema (e.g., how the manipulation of a mirror affects the reflection of light) may be more easily recalled if there is a cue to that representation in the form of a relevant gesture. The video instruction may also help memory for the gestures by activating the representational motor system and creating recall cues.

It has been argued that high spatial ability is related to more accurate ability to mentally animate, which is a key process in developing conceptual schemas of physics. Hegarty and Sims (1994) explained that both people with low and high spatial ability use the same process to develop mental animations. A mental animation is developed by first breaking down the information into causal links. For example, in the optics lesson, a mental animation could be developed for the concept of reflection that involves the causal links of moving a lens into a beam of light, rotating the lens, and determining the subsequent angle of light reflection. Each component of this process has a causal link to the next component. The mental animation process then is to animate these causal links by inferring the cause and effect movements that would occur. In a series of experiments, Hegarty and Sims (1994) found that those with low spatial ability were less accurate at inferring the links between the causal components of a mental animation. They argue this is due to lower working memory capacity because those with low spatial ability did worse on the later components in the causal chain, which would require more information to be held in working memory. People with high spatial ability, on the other hand, were better able to hold onto information in working memory so they did not suffer in mental animation accuracy. It follows that those with high spatial ability do better from the good instruction condition (i.e., Natural Video) because their spatial ability frees up working memory resources to focus on the mental animation component of the lesson and less on extraneous factors (i.e., arbitrary gestures or inability to recall text instructions), while those with low spatial ability are overburdened in their ability to process the conceptual information in working memory.

Similarly, the Arbitrary Gestures with Text Instruction condition was particularly bad for those with low spatial ability. This condition may have imposed working memory burdens that participants with low spatial ability were not able to overcome, while those with high spatial ability performed similar to those with high spatial ability in other conditions, which would support the spatial ability-as-compensator hypothesis (Mayer & Sims, 1994). The spatial ability-as-compensator hypothesis states that those with lower spatial ability will suffer from poor instruction, while those with higher spatial ability will be compensated by their spatial ability when instruction is worse. This condition could be poor instruction because the arbitrary gesture interactions did not relate to the learning material, so the interactions could distract from the conceptual information of the lesson. This theory might also have support from the open-ended responses of participants in the arbitrary conditions who stated they were too focused on the gestures to pay attention to the optics lesson. The burden of distracting gestures on mental processing could have been exacerbated by lower spatial ability, because those with low spatial ability are less able to retain mental animation information and the addition of extraneous load from arbitrary gestures disproportionately affected participants with low spatial ability. To determine the accuracy of the theoretical explanations that suggest learning is contingent on amount of cognitive load experienced during the computer lesson (i.e., spatial ability-as-enhancer and spatial ability-as-compensator), the instructional efficiency analysis later in the chapter combines cognitive load with learning.

Additionally, the text-based instructions of the interactions may have been worse than video instructions because text or verbal instructions are not as effective for teaching human motion tasks as video instructions that depict a human performing the action (Alexander, 2013). Observational learning of a human movement task in the form of a video tutorial may be better than a text-based tutorial because it activates the motor system of the learner, thereby priming the learner to conduct the same action (Van Gog, Paas, Marcus, Ayres, & Sweller, 2009). Without this benefit, and with the added burden of arbitrary gestures, those with lower spatial ability learned less than those with higher spatial ability who were compensated by better mental animation processing.

Cognitive Load

To answer the question of what interactions and instructions are best for gesture-based computer lessons, the next set of analyses tested how cognitive load was affected by these factors. Cognitive load was measured using a mental effort rating scale in which participants responded with their perceived level of mental effort on the computer lesson. Ratings of mental effort were compared for both types of Gesture interaction (Natural and Arbitrary) and Instruction (Video or Text) with spatial ability included as a potential predictor of cognitive load. Spatial ability did not account for a significant portion of the differences in mental effort ratings. At first glance, the lack of significance for spatial ability may seem to contradict the explanations of the spatial ability-as-enhancer and spatial ability-as-compensator hypotheses described in the previous subsection; however, spatial ability is a key factor in the analysis that tests cognitive load and learning together to determine instructional efficiency. Although spatial ability overall was not a predictor of cognitive load, it is not accurate to conclude that spatial ability did not affect cognitive load in the context of instructional efficiency. Spatial ability did interact with condition when assessed in conjunction with learning, and this analysis is discussed in the following subsection on instructional efficiency.

Cognitive load was explained mostly by the type of gesture-based interaction the participants used to complete the computer lesson. Those using natural gestures felt 25% less mental effort on the computer lesson than those using arbitrary gestures. This supported the hypothesis that when gestures are more closely mapped to the conceptual material in the lesson, less effort is required to use the gestures than when gestures are arbitrary. It could be that the arbitrary gestures increased the cognitive load felt by the participants by adding extraneous load to the learner. Arbitrary gestures increase extraneous load in that they are additional pieces of information that must be held in working memory during the computer lesson that do not directly relate to the conceptual material to be learned. Natural gestures that relate to the learning material could be considered germane load in that they aid in the understanding of the conceptual material by supporting the mental animation process. When the participant performs a gesture that is naturally-mapped to the conceptual material, they are physically performing the mental animation process of inferring motion between causal links.

Type of instruction did not affect cognitive load, contrary to the predicted hypothesis. It was predicted that video-based instructions of the gestures would lead to less cognitive load during the computer lesson than text-based instructions, because videos of a human performing the actions could activate the motor system and prime learners for the same action. Video instructions, unlike text-based instructions, have been shown to help learners understand hard to imagine tasks (such as the movement of a novel gesture), encourage multimodal processing, and recall concepts (Alexander, 2013). On the other hand, Ayres and Paas (2007) suggested that animated instructions (e.g., video tutorials) could increase extraneous load by creating a distraction from the lesson by requiring the learner to search for relevant information. Text-based instructions may be less likely to divert WM resources away from processing the lesson because they do not distract from subtle information, or text may have cleared up ambiguity that was not obvious in the videos. For example, the text states to start with the hand above the head for the action "down." If the hand did not start above the head for this gesture, it was not recognized by the motion tracker. That is, if the participant started with their hand slightly below their head, the motion tracker did not initiate the "down" command. Participants in the video condition saw a video of an actor performing the "down" gesture by starting with his hand above his head; however, participants may not have noticed that the hand was starting from a specific location without it being directly stated. Directly stating the starting point in the text condition may have avoided confusion.

Instructional Efficiency

The analyses culminated in the test of instructional efficiency, which utilizes standardized learning gain scores and mental effort ratings to create an estimate the efficiency of an instructional technique. Sweller, van Merriënboer, and Paas (1998) explained that instructional efficiency is used as a relative measure of cognitive load because it is difficult to measure the three types of cognitive load directly. The instructional efficiency measure has the benefit of allowing comparison of instructional techniques not only on how much learning occurred, but also the mental effort cost of learning in each factor. The first analysis tested whether differences in instructional efficiency existed based on Gesture type (Natural or Arbitrary) or Instruction type (Video or Text), as well as spatial ability. Spatial ability predicted instructional efficiency, which was expected because spatial ability was the main predictor of overall learning gain, so it should still be significant when extent of cognitive load was integrated with the learning gain measure. Type of gesture interaction in the computer system explained most of the differences in instructional efficiency of conditions. The conditions that used natural gestures had much higher instructional efficiency than either of the conditions using arbitrary gestures. This result, like the results in the previous section, lends evidence to the hypothesized theories in which naturally-mapped interactions benefit conceptual learning through more efficient instruction (i.e., more learning with less mental effort). Within each gesture type, there was very little difference between video and text instruction conditions on instructional efficiency, just as the type of instruction did not predict overall learning gains nor cognitive load. The type of gesture and type of instruction did not interact.

Because spatial ability was a significant factor in instructional efficiency of conditions, another follow-up analysis was conducted on the two manipulated variables, Gesture type and Instruction type, with the added variable of High and Low Spatial Ability. Spatial ability and type of Gesture were both significantly different between their respective levels, but they did not interact. Participants with high spatial ability had higher instructional efficiency than those with low spatial ability in every condition, and natural gestures were also more instructionally efficient than arbitrary gestures. Instruction types did not differ, again, and there were no other interactions among the factors.

It is important to revisit the spatial ability-as-enhancer and ability-as-compensator hypotheses that seemed to be supported by overall learning gain. When the amount of mental effort during learning was taken into consideration, there was evidence of a spatial ability-as-compensator effect for the poor learning condition, Arbitrary Text, such that people with low spatial ability had disproportionately worse instructional efficiency in this condition. It seems that, as explained in the learning gain analysis, those with high spatial ability are able to compensate for poor instruction. Additionally, there is evidence for the spatial ability-as-enhancer hypothesis in that the natural gesture with video instructions conditions was the most efficient for those with high spatial ability, but not for low spatial ability; although, the difference between video and text instruction in the natural gesture condition did not differ significantly, likely because there is a smaller range of standardized instructional efficiency scores.

In summary, instructional efficiency can be increased using natural gesture-based interactions, regardless of how the gestures were instructed and the learner's spatial ability. This finding supports the hypotheses that natural gesturing produces better conceptual learning by reducing cognitive load on the learner. Instructional efficiency can be enhanced when video tutorials are combined with natural gestures to produce an additive effect for those with high spatial ability. On the other hand, learners with low spatial suffer more from poor instructional design in the form of combining text-based instructions with arbitrary gestures.

Usability Analyses

In addition to answering the main research questions regarding which instructional techniques were best for a gesture-based conceptual computer lesson, additional analyses were conducted related to the usability of the system depending on condition. These analyses contribute to the broad investigation of gesture-based computer interactions and appropriate ways in which interactions can be instructed for educational and training systems. The following variables were included for both theoretical and applied implications for instructional design.

Time in Lesson

The amount of time to complete the computer lesson can be considered a secondary measure of instructional efficiency. The average time participants took to complete the computer lesson during the experiment was approximately 11 minutes. An analysis incorporating both manipulated variables (i.e., Gesture type and Instruction type) and spatial ability found that there was an interaction between the Gesture and Instruction type. Those using arbitrary gestures to interact with the computer system were slowest when instructions were text-based, and the arbitrary gesture conditions were slower than both natural gesture conditions by approximately two minutes. Spatial ability did not predict the amount of time participants took to complete the computer lesson.

System Usability

Perceived usability of the computer lesson was measured using the System Usability Scale. The analysis on usability scores was conducted with Gesture and Instruction types as well as spatial ability. Spatial ability did not predict perceived usability of any of the conditions in the computer lesson. There was an effect for the type of gesture interactions, such that natural gestures were rated almost a full point higher in usability (on a 7-point scale) than arbitrary gestures. This finding that natural gestures were perceived as having higher usability corresponds with the instructional efficiency finding that natural gestures led to better learning with lower mental effort, because there should be in inverse relationship between higher usability and lower cognitive load. There were no other effects or interactions that explained perceived usability.

Presence

Presence, or the feeling of "being there" in a virtual environment, was measured after the computer lesson using a presence questionnaire with four subscales: Involvement, Adaptation/Immersion, Sensory Fidelity, and Interface Quality. Although sense of presence *per se* was not a construct of interest in the current experiment, the subscales did include questions that were highly related to the study, such as asking how natural the interactions with the environment seemed. For all of the subscales except Sensory Fidelity, there was an effect for type of gesture interaction, such that natural gestures had higher "presence" ratings than arbitrary gestures. Natural gesture interactions were seen as inducing a higher sense of control in the computer lesson, more immersion, and a better interface quality.

Individual Differences

There were a few additional analyses related to individual differences that were potential predictors of the main outcome variables. Like the spatial ability analyses presented above, these subject variables were tested to rule out other explanations for results beyond the manipulated variables. These variables were not directly related to the research questions, but provide context for the results in relation to variables of interest in the literature.

Video Game Experience

The individual difference of video game experience was included in analyses as a potential predictor of learning and cognitive load, because video game experience could influence how a participant performs the computer task. Video game experience was approximated by asking participants how many hours a week they play video games, because hours a week playing video games has been correlated with both video game self-efficacy and comfort with video gaming (Procci, James, & Bowers, 2013). Number of hours a week playing video games was not directly related to either learning or cognitive load. Video game experience was also not a predictor of these outcome variables when included in statistical models with the manipulated variables (i.e., Gesture and Instruction type).

Gesture Production and Perception

Another individual difference that could influence performance in the computer lesson is the participant's production and perception of gesturing. For example, if a participant is less likely to produce gestures, he or she may be less likely to benefit from a gesture-based interface. A gesture survey was used to determine different dimensions of a participant's predisposition to gesture, with four subscales: Perception, Production, Social Production, and Social Perception. Of these four subscales, only the Production dimension was directly related to learning from the computer lesson, such that the more a learner tends to produce gestures in life, the more he/she learned from the computer lesson. None of the subscales was related directly to cognitive load. Because the Production subscale was related to learning, it was included in an analysis with the manipulated variables (Gesture and Instruction type) and the other significant predictor of learning, spatial ability. Even after an individual's spatial ability was accounted for, the individual difference of gesture production was still a marginally significant predictor of learning. Gesture production may be an important individual difference for gesture-based computer interactions and should be included in future studies that utilize gesture-based NUIs.

Revisiting the Research Questions

The set of studies presented here were conducted to determine whether type of gesture-based interaction, or instruction of those gestures, affects the learner on a conceptual lesson in a computer environment. The first research question was whether more naturally-mapped gestural interactions were better for learning from a computer lesson than arbitrarily-mapped gestures.

The results of the experiment indicated that when learning and cognitive load are combined in an instructional efficiency measure, natural gesturing is a much more efficient instructional technique than arbitrary gesturing. Based on these results that showed natural gestures were better for learning efficiently and other results that indicated natural gesturing was faster, had higher usability, and more presence than arbitrary gesturing, the results confirmed that natural gesture-based interactions are better for learners in a computer lesson than arbitrary gestures.

The second research question was whether types of instruction for the gesture-based interactions could influence the computer lesson. To test whether the effect of novel gesture-based interactions depends on how they are taught, the way the gestures were instructed was varied in the main experiment by using either video- or text-based tutorials. Results indicated that the type of instruction for the gesture-based interactions did not interact with the type of gesture overall, such that the detriment of arbitrary gestures was not overcome by instructions; however, when a participant's spatial ability was taken into account, the combination of gesture type and instruction type did seem important for learning. Those with low spatial ability had lower instructional efficiency with the combination of arbitrary gestures and text instructions, while those with high spatial ability were benefited by natural gestures with video instructions. Based on the results, the research question can be answered: video- or text-based instructions of the gesture-based interactions did not influence the computer lesson overall, but there could be a combinatorial influence of gesture type and instruction type depending on spatial ability of the learner.

In summary of the research question answers, naturally-mapped gesture interactions were better than arbitrary gestures for both conceptual learning and usability outcomes, and may especially benefit those with higher spatial ability when combined with video instructions.

Conversely, those with low spatial ability may be particularly disadvantaged when arbitrary gestures are taught using text-based instructions.

Summary of Theoretical Implications

Embodiment Theory

The results of the current experiment in which gesture-based interactions with the computer lesson were either naturally-mapped to the learning material or arbitrary, supported theories of Embodied Cognition. The embodiment theories described in the Introduction chapter can be summarized as ways in which physical interactions with one's environment affect mental processing and representations. Specifically, when actions are physically performed, they serve to activate the motor system, creating stronger memories through enactment and helping to develop schemas for the action (Barsalou, 2008; Engelkamp & Jahn, 2003; Hostetter & Alibali, 2008). When a gesture-based interaction is relevant to the learning material of a lesson, the gestures can be considered naturally mapped, or "enactive mapping" (Schwartz & Plass, 2014). Because the natural gestures consisted of enacting the learning material, the natural gestures contributed to more instructional efficiency in that more conceptual information was learned with less mental effort expended. This finding supports embodiment theories in that physically enacting relevant information to the conceptual lesson helped understand of the conceptual information.

It could be that natural gesturing created stronger memories or more accurate mental representations for the optics concepts that were not enacted in the arbitrary gesture conditions. When the learner used a naturally-mapped gesture to interact with the lens or mirror in the lesson, they were performing the movement directly related to the optics concept; for example, the learner would move his or her hand left to move the lens/mirror to the left. Then, the beam of light would reflect/refract depending on the angle and type of lens or mirror, so that the learner could visualize optics concepts by manipulating the lens/mirror and observing the result on the beam of light. Using natural gestures that corresponded with the movement of the lens/mirror could therefore create a physical encoding for the concept in that the gestures were helping the learner to make the link between the movement of the lens/mirror and the resulting effect on the beam of light. In contrast, the arbitrary gestures did not correspond directly with the movement, so performing a gesture would not be related to the optics concept. Because the arbitrary gesture did not match the conceptual information in the lesson (i.e., the gesture and the result on the lens/mirror were different movements), the arbitrary gestures may not help the learner visualize the movement of the lens/mirror as the gesture and optics concept would be encoded as two different movements. This would add the burden of encoding and recalling the arbitrary gesture movement in addition to the movement of the lens/mirror (i.e., two pieces of interacting information about movement), as opposed to the natural gestures that were the same movement as the lens/mirror (i.e., one piece of information about movement). In addition to the burden of additional information to encode and recall, arbitrary gestures may not help recall of the concept because the arbitrary gestures were not related to the movement of the lens/mirror and an additional link would need to be made between these pieces of information, instead of the gesture assisting in the associative link between action and result. Conversely, the natural gestures were creating stronger memories for the optics concepts by encoding the information in a way that can be more easily recalled because the encoding of the gesture can act as an additional cue for recalling the conceptual information, while the arbitrary gestures do not have this benefit. Recalling the natural gestures may activate the mental simulation for the optics concept because these actions are stored together in the sensorimotor system, while the arbitrary gestures may be stored as separate information that is not directly related to the optics concepts. The natural gestures may therefore produce more accurate recall of the conceptual information because of the strong associative link between the gesture and the resulting movement on the beam of light. The evidence from the current experiment supports the theories of Embodied Cognition that enacting the learning material leads to better for instructional efficiency in a conceptual learning computer lesson.

Cognitive Load Theory

In addition to implications for Embodied Cognition, there are theoretical implications for CLT from the current experiment. The main tenant of CLT is that a person's working memory has a limited capacity for processing new information, and cognitive load is the amount of information being processed by working memory at one time (Sweller, van Merriënboer, & Paas, 1998). Cognitive load consists of three types of load (i.e., intrinsic, extraneous, and germane) that, when combined, can overload working memory capacity and impede information processing (i.e., learning). The theory states that intrinsic load, which is the load associated with the difficulty of the material, cannot be changed, so the goal of instructional design is to reduce extraneous load due to factors not related to learning and foster germane load that helps schema development from new information. As described by Sweller et al. (1998), instruction can reduce extraneous load and increase germane load by directing attention to relevant information during a lesson, thereby helping learning.

There is evidence in the current experiment that the type of gesture-based interaction with a computer lesson are explained by CLT, although direct measurements of each type of cognitive load were not possible (see limitations subsection below). Natural gestures contributed to more instructional efficiency than arbitrary gestures, which is a relative measure of cognitive load.

These results can be explained by CLT in that the arbitrary gestures were not related to the conceptual information and increased extraneous load by adding information to be processed in working memory that was not related to the lesson. Natural gestures may have decreased extraneous load by directing attention to the relevant conceptual information, thus aiding processing of the concepts.

The results also support that the underlying cognitive mechanism leading to the difference in instructional efficiency may be due to the modality of encoding, as described by Baddeley's (2000) working memory model. The modality of the instructions was varied by presenting either video instructions (i.e., visual modality) or text-based instructions (i.e., verbal modality) for the gesture-based interactions. It was expected that the video instructions would be better than text instructions due to a modality effect in which seeing the gestures would prime the motor system for the actions and help participants to visualize hard to imagine material (i.e., human movement; Alexander, 2013). Presenting the gesture instructions as textual information in the tutorial would be processed as verbal information in the phonological loop. Because the conceptual information in the optics lesson was presented as narrated text, this lesson may also be processed as verbal information in the phonological loop. If both the gesture instructions and the conceptual information were processed in the phonological loop as verbal information, it may have overloaded mental processing in the verbal system and made for less efficient instruction. Conversely, presenting the gesture tutorial as videos would result in the gesture instructions (i.e., visual) processed separately from the narrated text of the optics lesson (i.e., verbal), thereby not overwhelming the phonological loop with too much simultaneous information to process. The type of instructions only mattered when combined with the type of gesture and when spatial ability was taken into consideration. Video instructions did help instructional efficiency when combined with natural gestures for learners with high spatial ability, while text instructions combined with arbitrary gestures was worse for instructional efficiency for low spatial individuals; so, there was a modality effect, but only under certain circumstances. The lack of strong support for a modality effect may be due to the strength of the manipulation, which is discussed in the limitations section below.

Summary of Empirical Implications

The set of studies developed for the current research also have empirical implications for the study of NUIs. Although previous research on gesture-based interactions has focused on what gestures the computer can recognize (Nielsen, et al., 2004; Shiratuddin & Wong, 2011), gesture design does not typically begin with a user-centered approach that takes into consideration what the user perceives is natural. The issue with studying natural gesturing is that the researcher must first confirm that the gestures are interpreted as natural. The methodology developed for the current research questions can be extended to other research questions investigating natural gesturing. The first study explored what natural gestures are spontaneously produced by participants to narrow down what gestures may be considered natural for each action. This methodology took a user-centered approach in that what a user would produce for a gesture-based computer command was considered prior to what is easiest for a computer to recognize. Participants were asked to perform a gesture they considered natural for each computer action, and gestures were recorded using a motion tracker for video, depth, and joint placement information. The gestures were then analyzed for converging features using the coding scheme outlined in Study 1. The most common gestures for each action were then chosen for the second part of the NUI development that asked a different group of participants to rate how natural they felt each gesture was for a computer action. Participants watched a short video of the gestures including those gestures that were performed in Study 1, and then they rated the naturalness of each gesture for a particular action on a scale from "Completely Natural" to "Completely Arbitrary." The ratings for the gesture videos were then analyzed for each action using Repeated Measures ANOVAs described in Study 2, and the gestures rated as most natural for each action were determined. The results of the second study confirmed that the gestures produced by the participants in the first study were rated as natural. This has empirical implications for the study of NUIs in that the user-centered approach to determining what gestures are natural resulted in gestures that were interpreted as natural by other users. After narrowing down the potential natural gestures for each action by the choosing the most common gestures that were produced and picking the gesture rated as most natural for each action, the natural gestures were tested using the motion tracker to confirm that the gestures were possible to implement in the NUI. To summarize, future studies of NUIs can determine the natural gestures to be implemented in the interface by using a two-step process: 1. Ask participants what they think is natural and 2. Ask a second group of participants if they interpret those gestures as natural.

Summary of Applied Implications

The results of the current set of studies can be applied to educational and training systems that incorporate a gesture-based NUI. The finding that more natural gestures are better for learning efficiency, cognitive load, and a variety of usability factors should encourage instructional designers and system engineers to keep the user in mind when developing the gesture-based interactions. The instructional efficiency and usability of gesture-based interactions were not greatly impacted by how they were instructed, so designers should not rely on tutorials to overcome the limitations of arbitrary interactions. For learners with higher spatial ability, there was an added benefit of natural gestures that were taught using video tutorials. Because natural gestures with video instructions led to the best learning for certain participants, and natural gestures overall were better regardless of instruction, instructional designers should consider using both natural gestures with video instructions. Natural gestures taught via video instructions may particularly benefit learners who may not be able to read text-based instructions (e.g., learners who speak a different language than that in which the system was developed) or for younger participants before they learn to read (e.g., early elementary education).

Guidelines for Application

Grandhi et al. (2010) proposed guidelines for developing NUI interfaces, and I extend these by outlining a more general methodology from which new interfaces can be developed. Based on the results of this experiment, below I list guidelines for implementing gesture-based interactions into education and training computer systems:

1. Use gestures that are natural in the sense that they correspond directly with the learning material. Gestures may be considered arbitrary if they do not reflect the real-world interaction with an object that they represent, or if they do not aid in the mental animation process associated with developing a schema for the learning material.
2. When possible, determine what gestures are considered natural for an interaction directly from user input instead of relying on what is easiest for the computer to recognize. The current set of studies used a two-phase process for developing the natural gestures by first asking participants to produce a natural gesture for an interaction and then having other participants rate the naturalness of the interaction. This process produced a vocabulary of natural gesture interactions that was better for the learning material and usability.

3. Use video-based instead of text-based tutorials to instruct the gestures, but do not assume that instructions can overcome the detriment of arbitrary gestures. Video tutorials were best in combination with natural gestures for certain users, but natural gestures were best for all conditions regardless of instruction type.

Limitations

Measurement of Cognitive Load

A limitation of the current study is that the theoretical explanations for results are often based on the amount of cognitive load in each of three components (i.e., intrinsic load, extraneous load, and germane load), but a direct measurement of each component does not exist in the literature (Brünken, Seufert, & Paas, 2010). Brünken and colleagues (2010) described the three ways in which cognitive load is typically measured as subjective, objective, and combined. The measure used in the current study is considered a subjective measurement of mental load, which is an approximate of overall cognitive load. This measure of cognitive load was used in the current study because, as Paas, Tuovinen, Tabbers, and van Gerven (2003) explained, "The scale's reliability and sensitivity and moreover its ease of use have made this scale, and variants of it, the most widespread measure of working memory load within CLT [Cognitive Load Theory] research (p. 68);" however, there are several limitations of this measure described by Brünken et al. (2010). First, there is an assumption of subjective measures that the participant can accurately determine their mental effort. The measure assumes that the number the participant reports on the scale for mental effort is an accurate reflection of their cognitive load and can be comparable to how other participants interpret the scale. With this assumption, there is a risk that the participant is not accurately able to identify a number corresponding to their mental effort and that interpretation to be equivalent to other participants. Also, the report of mental effort could relate to any of the three components of cognitive load. For example, while theory predicts that more natural gesturing would reduce extraneous load, measurement of cognitive load cannot distinguish a reduction in extraneous load with a reduction in germane load, which would be detrimental to learning.

Although there have been recent attempts at creating a measurement that distinguishes the three components of cognitive load (Leppink, Paas, van der Vleuten, van Gog, & van Merriënboer, 2013), there is not strong support for this measure, and subsequent research has proposed various modifications (Leppink & van den Heuval, 2015; Leppink, Paas, van Gog, van der Vleuten, & van Merriënboer, 2014). To address the limitations of using a single subjective measure of cognitive load, Brünken et al. (2010) explained that a way to determine how cognitive load is related to performance, and thus indirectly measure the different components of load, is to use combined measures of cognitive load. The instructional efficiency measure calculated in the current experiment is one such combined measure of cognitive load and performance (i.e., learning) that creates a relative measure of how these factors interact; however, the issue of no direct measurement of cognitive load still exists with this method.

Manipulation Strength

Another limitation of the current experiment could be the strength of the instruction manipulation. The manipulation of instruction was intended to be video- versus text-based tutorials that instructed the gesture-based interactions. It was anticipated that video instructions would be better than text instructions because videos may reduce cognitive load and activate motor mental representations (Alexander, 2013). This is because video-based instructions can help the learner visualize the movements of the gestures and store that information as visual information, which can be mentally simulated in the sensorimotor system; however, text-based instructions may be stored as verbal information, resulting in a mental simulation of verbal information instead of a visualization for the gesture. This hypothesis was largely not supported, except in combination with gesture type and depending on participant spatial ability. The lack of results with instruction may have been due to how the instruction conditions were manipulated. The instruction manipulation not have been strong enough because all groups were able to use a summary sheet of gestures during the experiment and practiced the gestures during the tutorial. These adjustments were made to the conditions based on pilot testing of the experiment in which participants were not able to begin the experiment if they had misconceptions about the gestures, or they were not able to finish the computer lesson if they forgot a gesture during the experiment. All conditions were told to perform the gestures at the end of the tutorial to confirm the gesture was correct, resulting in a pre-training effect (Mayer & Moreno, 2003). If the gesture was not correct, it would not be recognized by the motion tracker and the gesture type manipulation would be confounded due to participants using inconsistent gestures; however, the instruction manipulation may not have been strong enough in that every person could have received the benefit of an enactment effect, in that physically enacting the gestures encoded the actions in the physical modality in addition to the encoding from the video and text instructions. Because all conditions received additional instructional support in the form of memory aids (i.e., gesture reference sheets) and an enactment effect (i.e., performing the gestures), the difference between video and text instructions may have been attenuated such that the effect was only found in combination with the gesture type.

Future Directions

Future research could expand on the current study in several ways. First, the research question asking whether naturally-mapped gesture interactions were better than arbitrary gestures was answered with support for natural gestures. The task in the current study was a conceptual optics lesson, and future research could continue this work with other domains (e.g., medical education, maintenance training, etc.). Perhaps natural gestures are appropriate for some kinds of knowledge or skill acquisition (e.g., conceptual information or procedural tasks), but not all types of knowledge (e.g., semantic).

Additionally, future research could incorporate features to make gesture-based interactions even more natural. The gestures included in this study were rated as natural by participants, but they were not dynamic in the sense that as the participant moved the on-screen object was not manipulated in real-time. If pantomimic gestures were closer to a 1:1 mapping of user-movement to computer-movement, the gestures may be perceived as even more natural than distinct gestures. Gestures could also be made more natural by including more fine-grained movements. The gesture-based interactions in the current study were gross movements that could be recognized by the low-resolution motion tracker (although development of gestures was user-based and not based solely on computer recognition), but better motion trackers may be able to recognize even more fine-grained gestures than those used in this experiment.

To expand on the second research question that assessed how much instructions of gesture-based interactions affect the computer lesson, future research could include variations of instructions to explore this question further. Instructions and tutorials could be presented in other media, such as spoken text or only physical enactment to address the limitation of the current study that necessitated some enactment. Alternatively, instructions could be presented in multiple modalities versus single modalities to address this same limitation and expand on the modality effect literature.

APPENDIX A

DEMOGRAPHIC SURVEY

DEMOGRAPHICS

1. Age: __________
2. Sex: __________
3. Major: ________________________________
4. Ethnicity (Please select all that apply):

    American Indian/ Alaskan Native

    Arabian/Middle Eastern

    Asian/Pacific Islander

    Black/African-American

    Hispanic/Latino

    White/Caucasian

    Other

    Prefer not to respond
5. Are you colorblind? Yes No
6. Dominant Handedness: Right Left Ambidextrous
7. Do you have normal or corrected vision (i.e., glasses, contacts)? Yes No
8. Highest level of education completed:

    Less than High School

    High School

    Some College

    Bachelor's Degree

    Advanced Degree
9. Please enter the typical number of hours per week that you use a computer: _________
10. Do you own a personal computer?

    Yes

    No
11. On what platform do you usually play video games?

    Game console

    Computer

    Phone/Mobile Device

    N/A

12. How often do you play computer games?
    - Daily
    - Weekly
    - Monthly
    - Less than once a month
    - Never
13. How often do you play video games (run on a console, not a computer)?
    - Daily
    - Weekly
    - Monthly
    - Less than once a month
    - Never
14. Please enter the typical number of hours per week that you play video games: ________
15. How long have you been playing video games?
    - N/A
    - 6 months
    - 1 year
    - 2-5 years
    - 5-10 years
    - 10 or more years
16. Please rate your skill at playing video games:
    - Bad
    - Poor
    - Average
    - Better than average
    - Good
17. Do you have any experience using motion-capture systems (e.g., Microsoft Kinect)? If so, please explain your experience and name the system.

    ____
    ____
    ____
    ____

18. What are your Top 3 (in order) video game categories/genres that you enjoy playing? (Choose from the list below, or add your own).

    1) ______________________________

    2) ______________________________

    3) ______________________________

Video Game Genres (Question 18)
Action
Fighting
First-person shooter
Role-playing
Massively Multiplayer Online Games
Simulators
Flight
Racing
Sports
Military
Space
Strategy
Strategy wargames
Real-time and turn-based strategy games
Real-time tactical and turn-based tactical
God games
Economic simulation games
City-building games
Adventure
Arcade
Educational
Maze
Music
Pinball
Platform
Puzzle
Stealth
Survival/horror
Vehicular combat
Other (please specify)

APPENDIX B

KNOWLEDGE OF OPTICS TEST

Knowledge of Optics Pre- and Post-Test

1. What type of lens is a magnifying glass? ____________________________
2. What is the bending of light rays as the rays pass through a substance called? ____________________________
3. What type of mirror makes objects appear smaller, but the area of view larger? ____________________________
4. Mirrors ____________________________ light rays to make an image.
5. A(n) ____________________________ mirror is like the side mirrors on a car ("Objects are closer than they appear")
6. A typical mirror you look in at home or in a restroom is a(n) ____________________________ mirror.
7. Any smooth surface that reflects light to form an image is a(n) ____________________________
8. The place at which light rays converge is the ____________________________
9. ____________________________ mirrors diverge light.
10. ____________________________ lenses diverge light.
11. A ray of light that approaches a mirror is a(n) ____________________________ ray.
12. A ray of light that reflects off a mirror is a(n) ____________________________ ray.
13. The perpendicular line that can be drawn from a mirror that divides an approaching ray from a reflected ray
14. What object reflects light and curves inward? ____________________________
15. What kind of mirror is used in a headlight, flashlight, or spotlight to create a beam of light? ____________________________
16. A convex mirror will always produce an image that is
    A) Real, upside down, smaller
    B) Virtual, upright, same size
    C) Virtual, upright, smaller
    D) Virtual, upright, larger
    E) I don't know

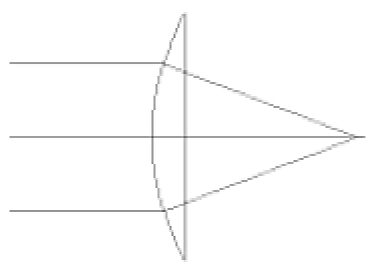
17. Are you reading all questions and answering honestly? (Attention Check)

18. A concave lens will always produce a(n) __________ image.
    A) Virtual, upright, smaller
    B) Real, inverted, smaller
    C) Real, inverted, larger
    D) Virtual, upright, larger
    E) I don't know

19. What is a refracting object that is thicker in the center than it is at the edges?
    ____

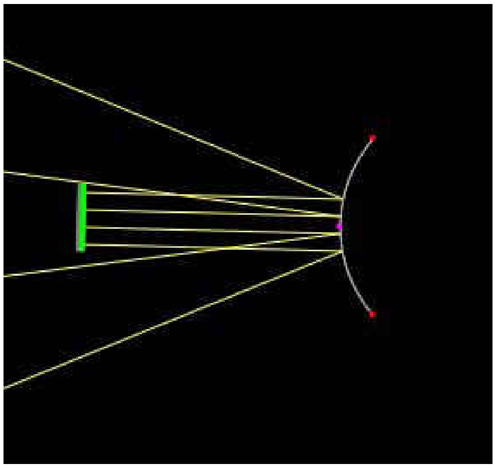
20. What is required for your eye to see an object?
    A) A mirror
    B) Air
    C) Light coming from an object
    D) A telescope
    E) I don't know

21. Which properties describe images formed by the lens in the figure?



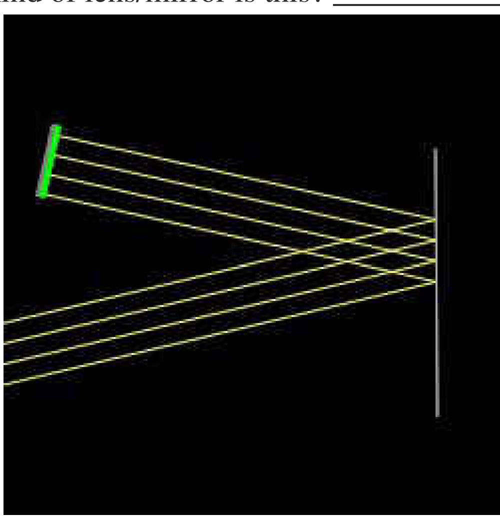
    A) Upright, larger than the object
    B) Upright, smaller than the object
    C) Upside down, larger than the object
    D) Upside down, smaller than the object
    E) I don't know

22. Which of the following describes images formed by the mirror when the object is in front of the focal point?

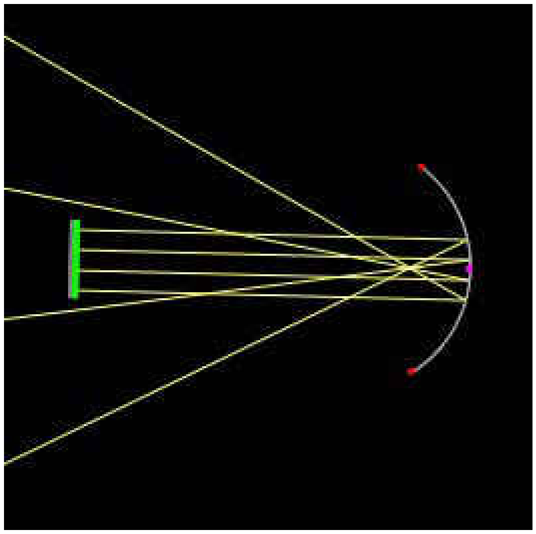


A) Inverted, larger than object
B) Inverted, smaller than object
C) Upright, smaller than object
D) Upright, larger than object
E) I don't know

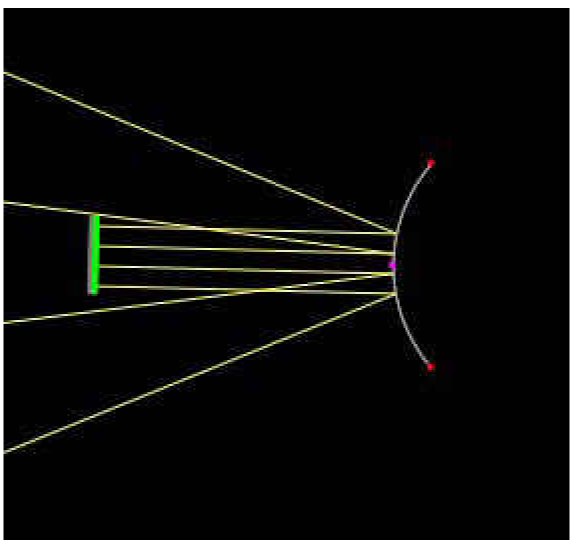
23. What kind of lens/mirror is this? ______________________________



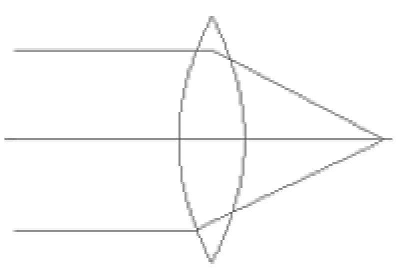
24. What kind of lens/mirror is this? ___________________________



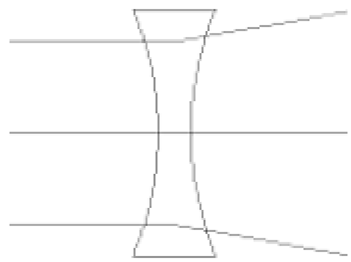
25. What kind of lens/mirror is this? ___________________________



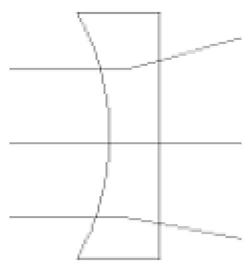
26. What kind of lens/mirror is this? ___________________________



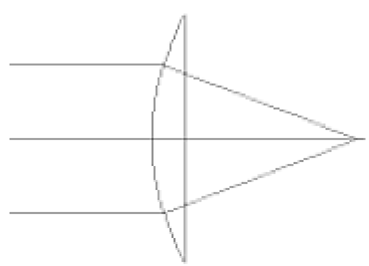
27. What kind of lens/mirror is this? ____________________________



28. What kind of lens/mirror is this? ____________________________



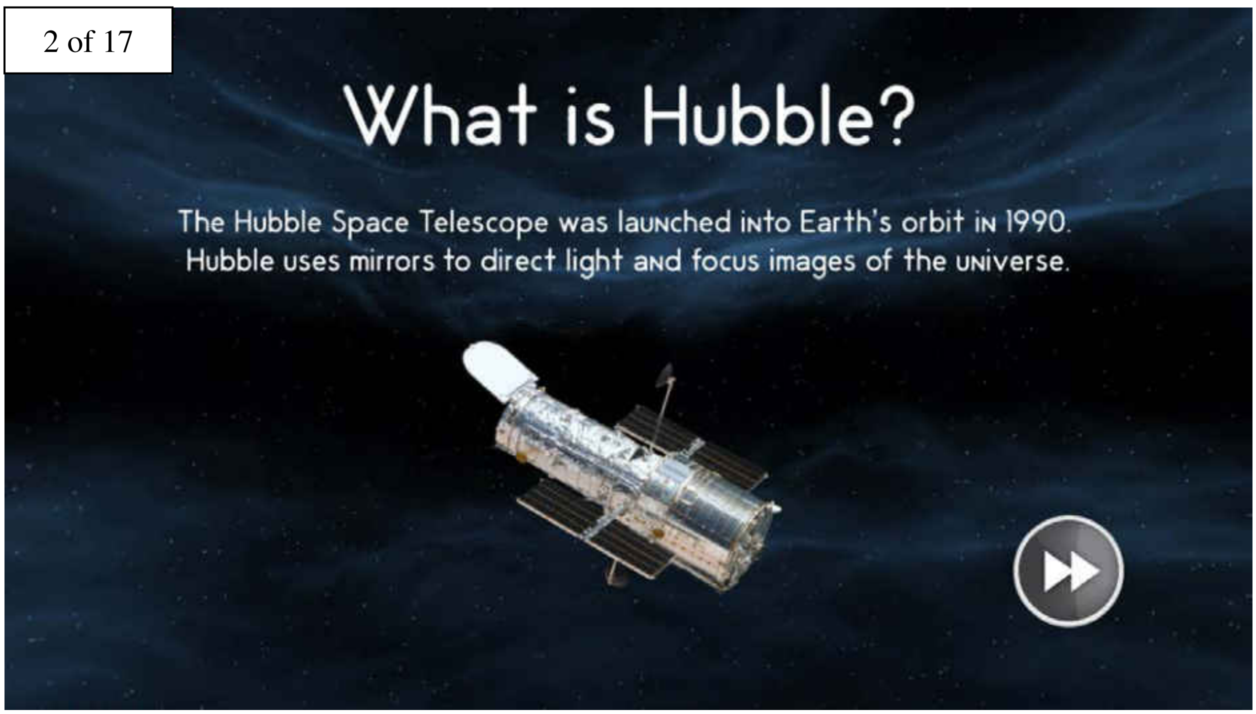
29. What kind of lens/mirror is this? ____________________________



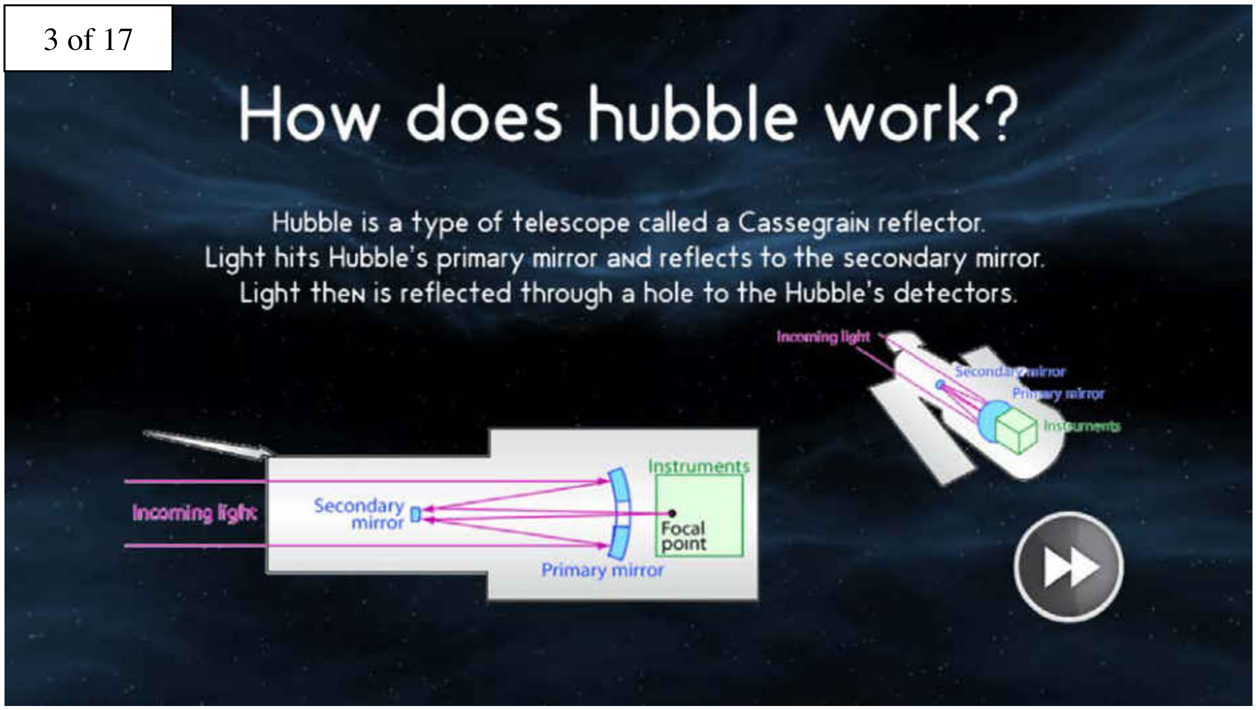
APPENDIX C

EXPERIMENT TESTBED SCREENSHOTS


HUBBLE
NEEDS
GLASSES


What is Hubble?
The Hubble Space Telescope was launched into Earth's orbit in 1990.
Hubble uses mirrors to direct light and focus images of the universe.

3 of 17


How does hubble work?
Hubble is a type of telescope called a Cassegrain reflector.
Light hits Hubble's primary mirror and reflects to the secondary mirror.
Light then is reflected through a hole to the Hubble's detectors.
Incoming light
Secondary mirror
Primary mirror
Instruments
Focal point

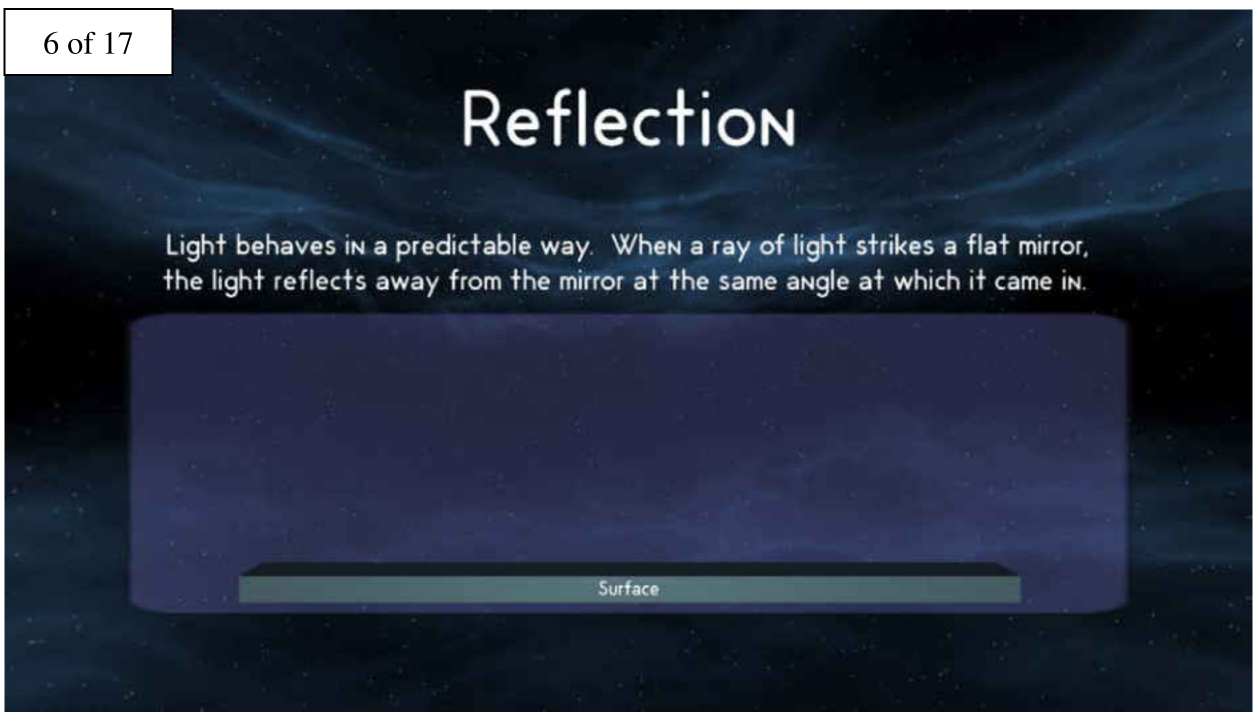
4 of 17


What Went Wrong?
...but the first images Hubble sent back to Earth were blurry!
One of Hubble's mirrors was too flat, which scattered light
and made a fuzzy image.

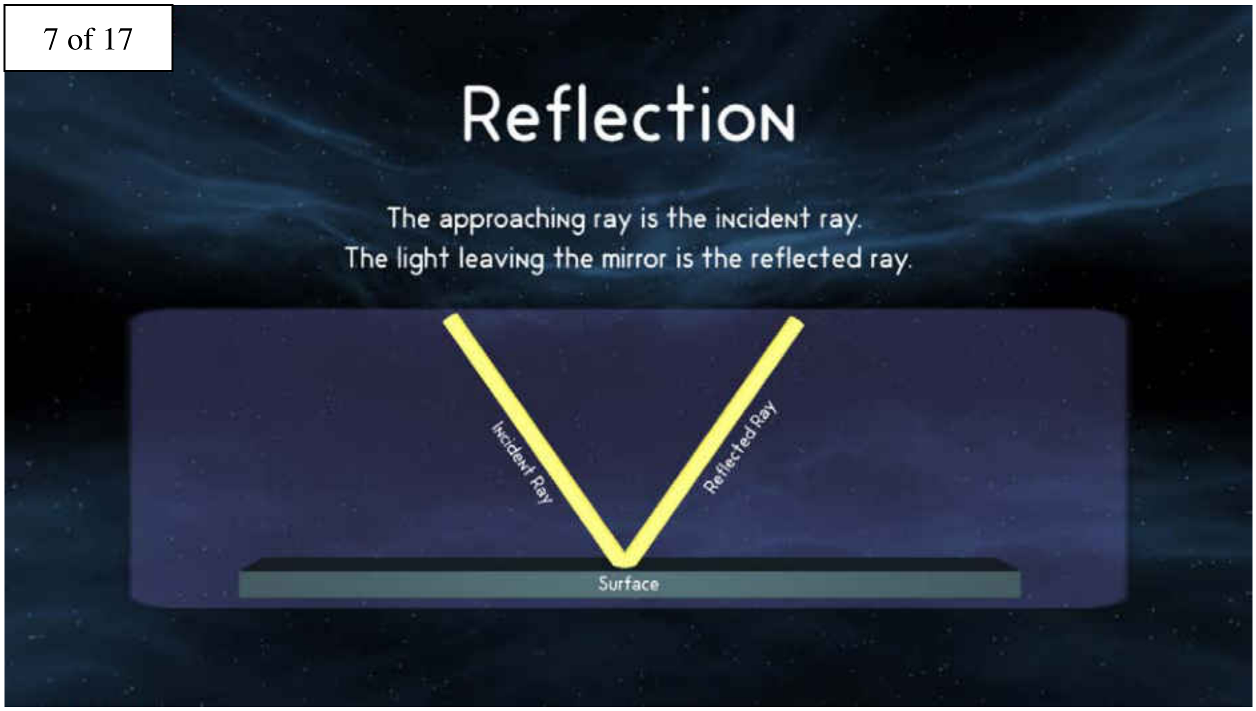
5 of 17


How was Hubble fixed?
Astronauts inserted corrective mirrors to direct light and make images clear.
Now let's take a closer look at how mirrors and lenses work...

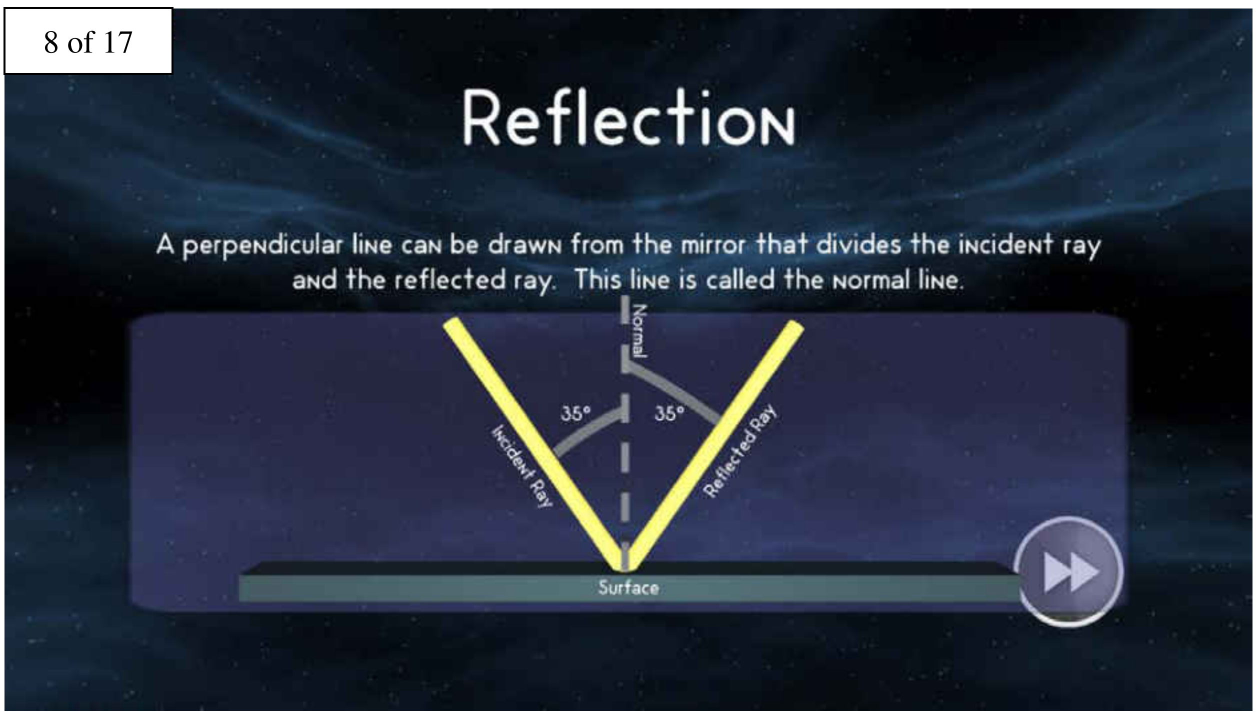
6 of 17


Reflection
Light behaves in a predictable way. When a ray of light strikes a flat mirror, the light reflects away from the mirror at the same angle at which it came in.
Surface

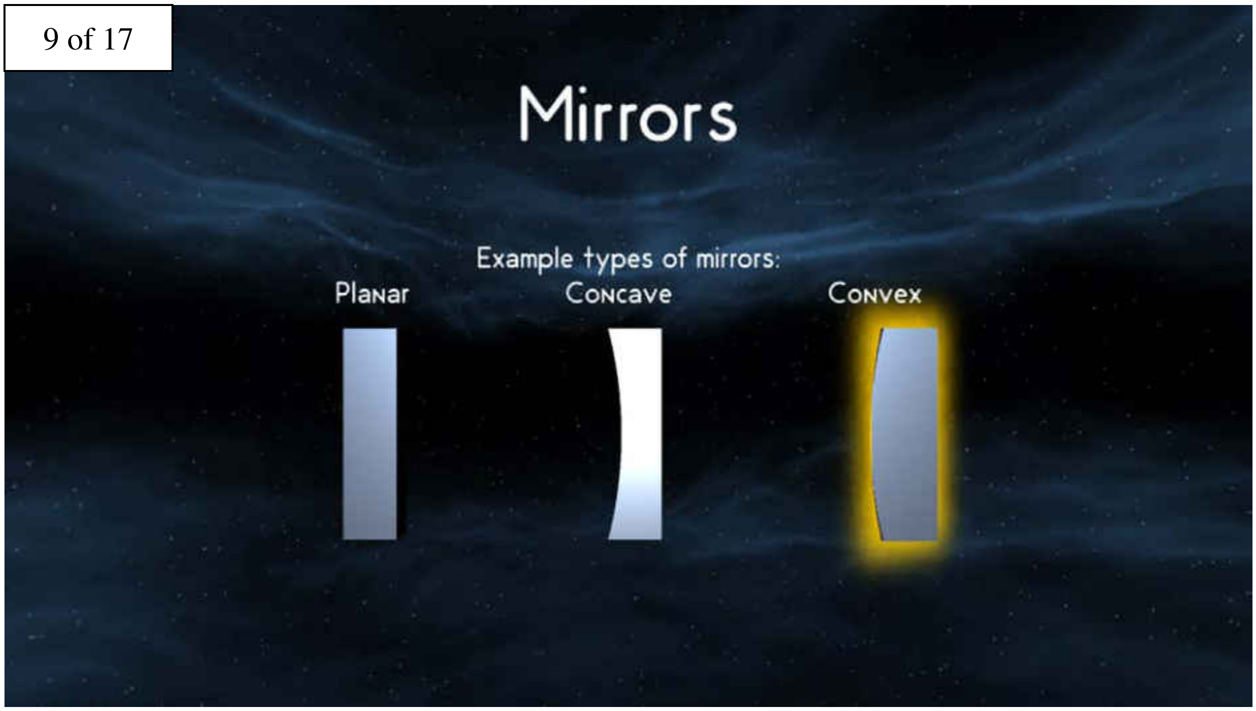
7 of 17


Reflection
The approaching ray is the incident ray.
The light leaving the mirror is the reflected ray.
Incident Ray
Reflected Ray
Surface

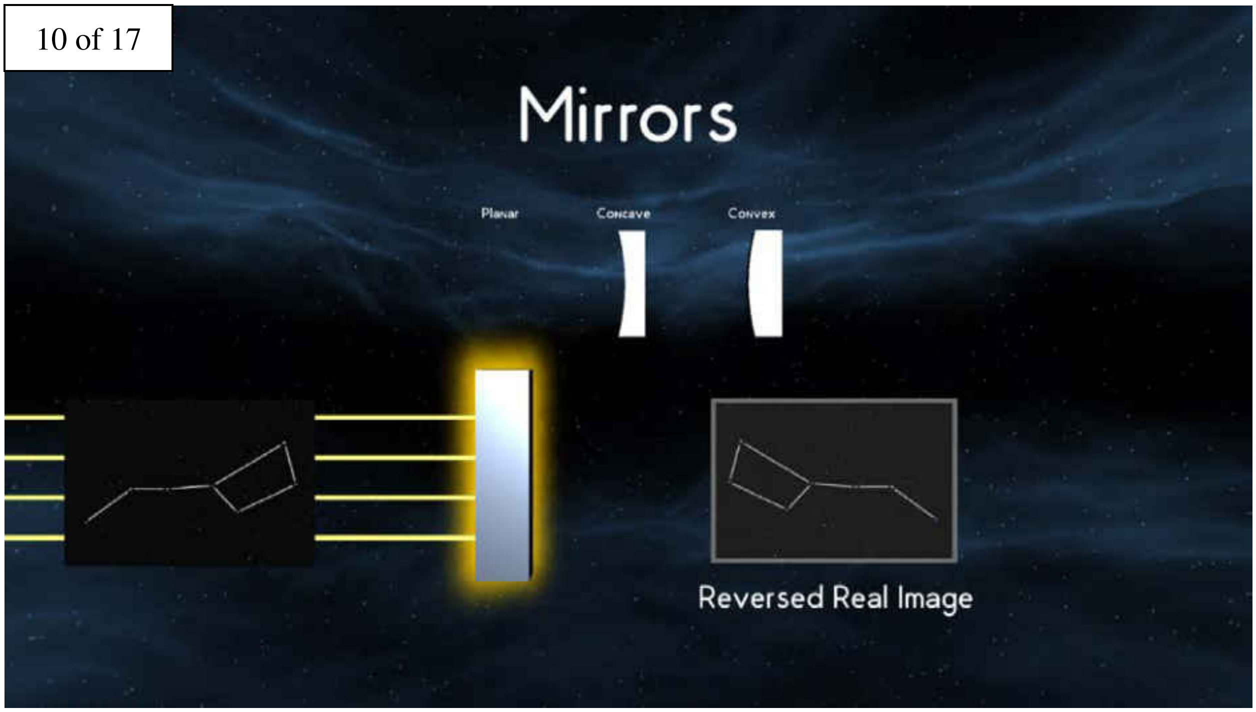
8 of 17


Reflection
A perpendicular line can be drawn from the mirror that divides the incident ray and the reflected ray. This line is called the normal line.
Normal
35°
35°
Incident Ray
Reflected Ray
Surface

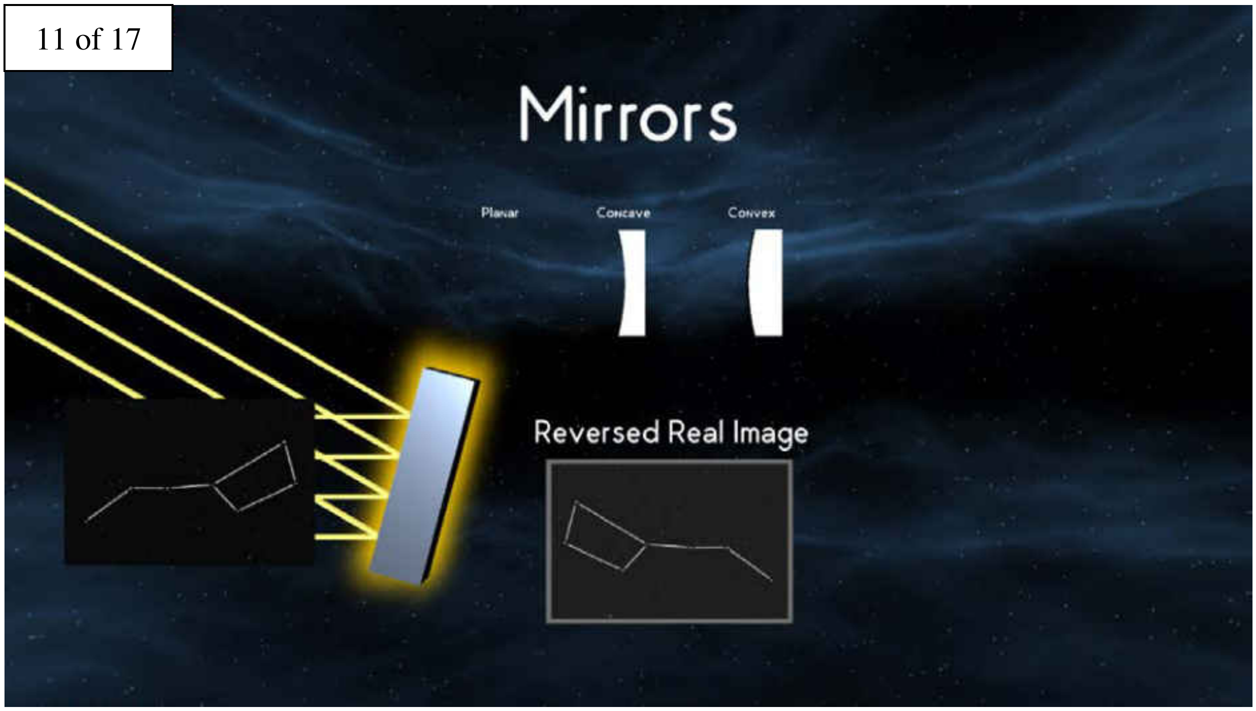
9 of 17



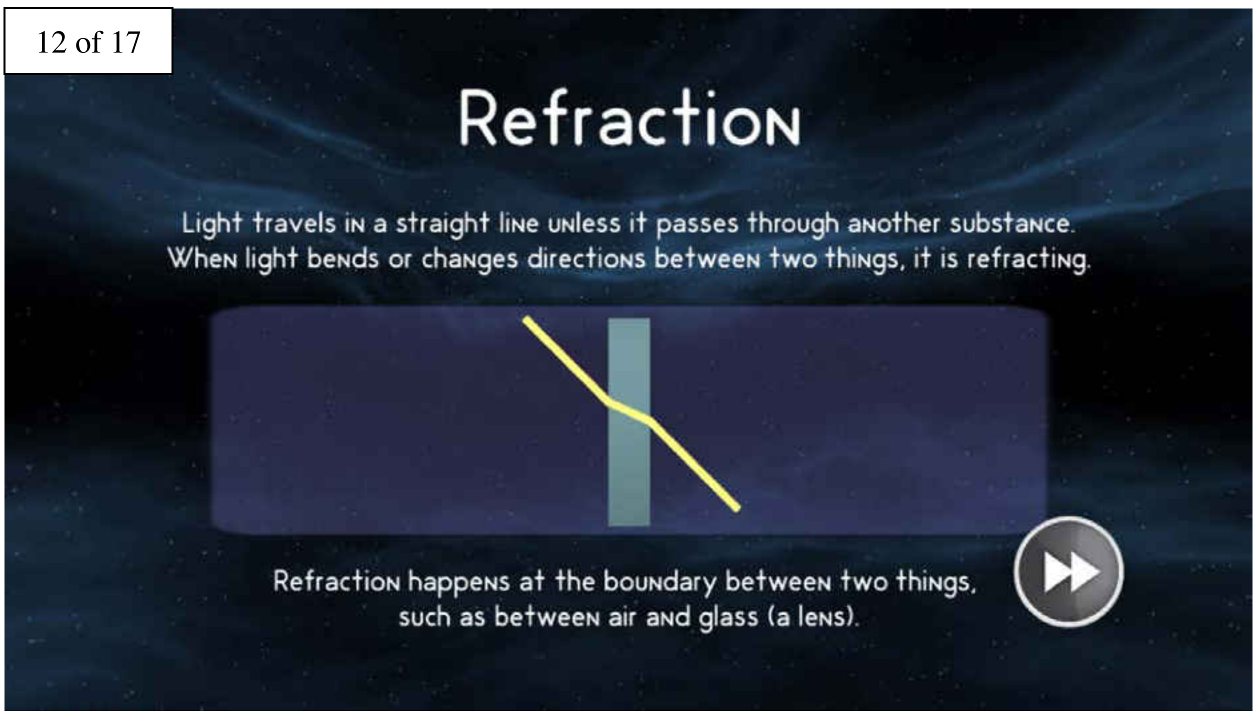
10 of 17



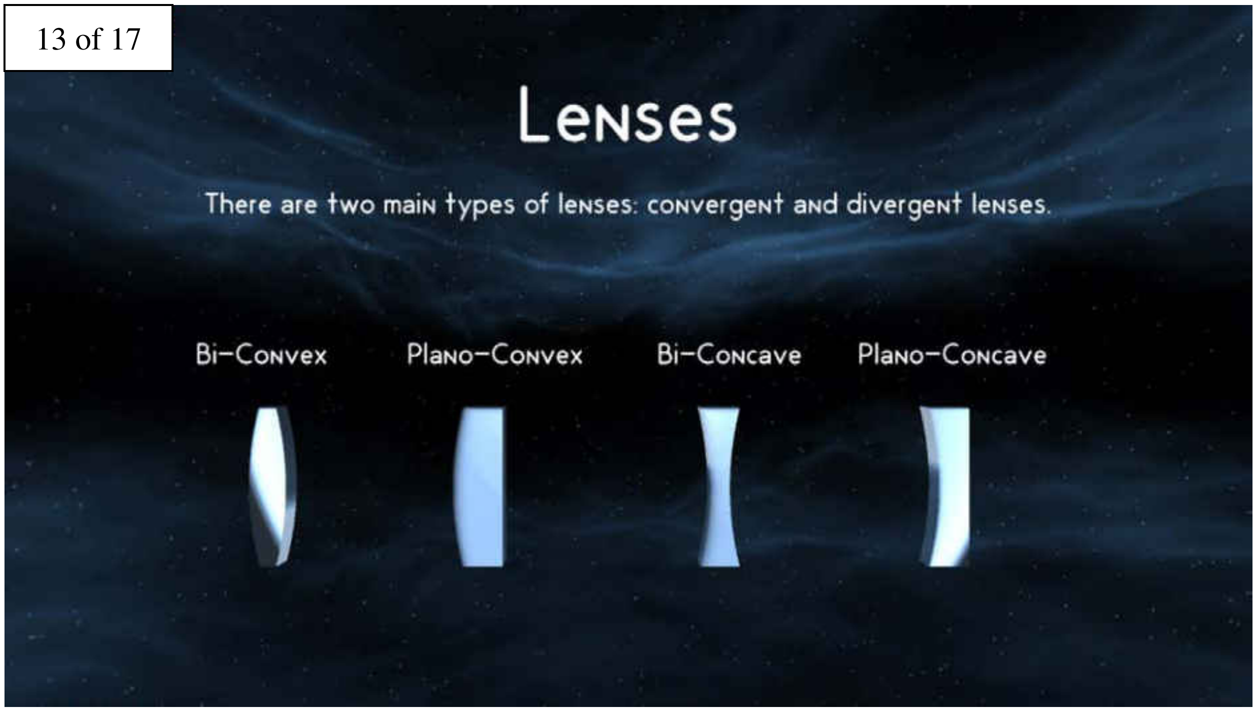
11 of 17



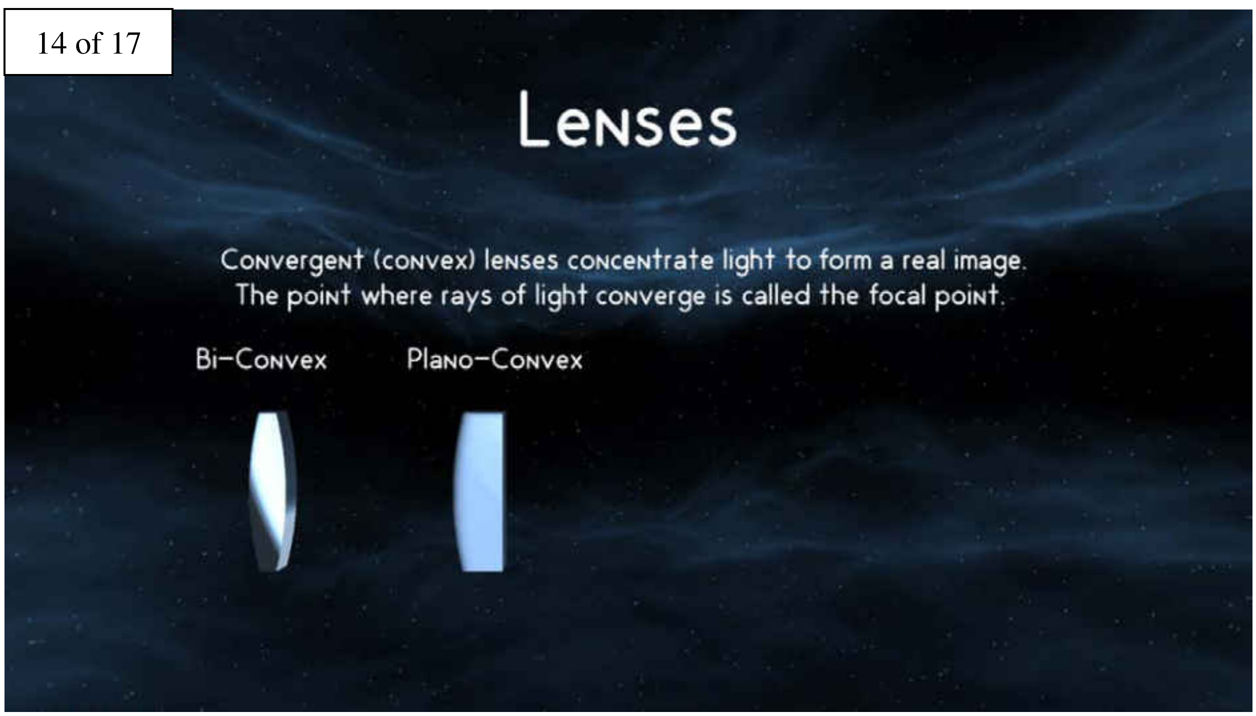
12 of 17

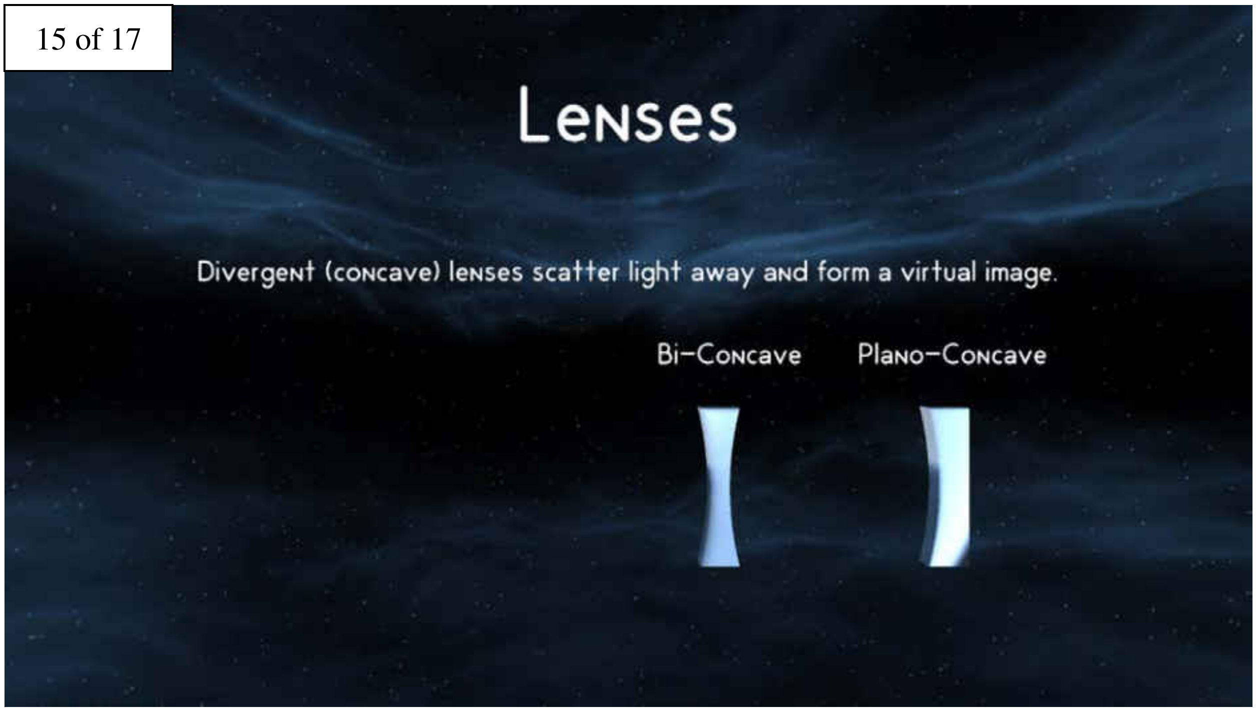


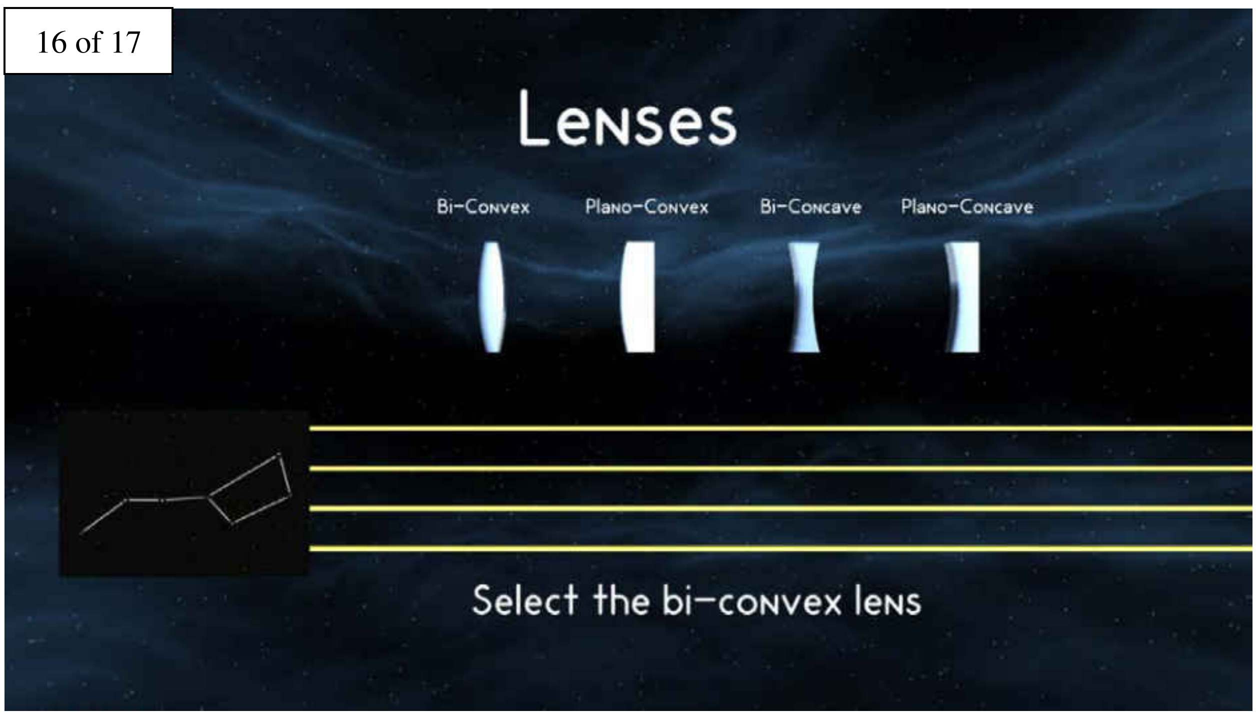
13 of 17

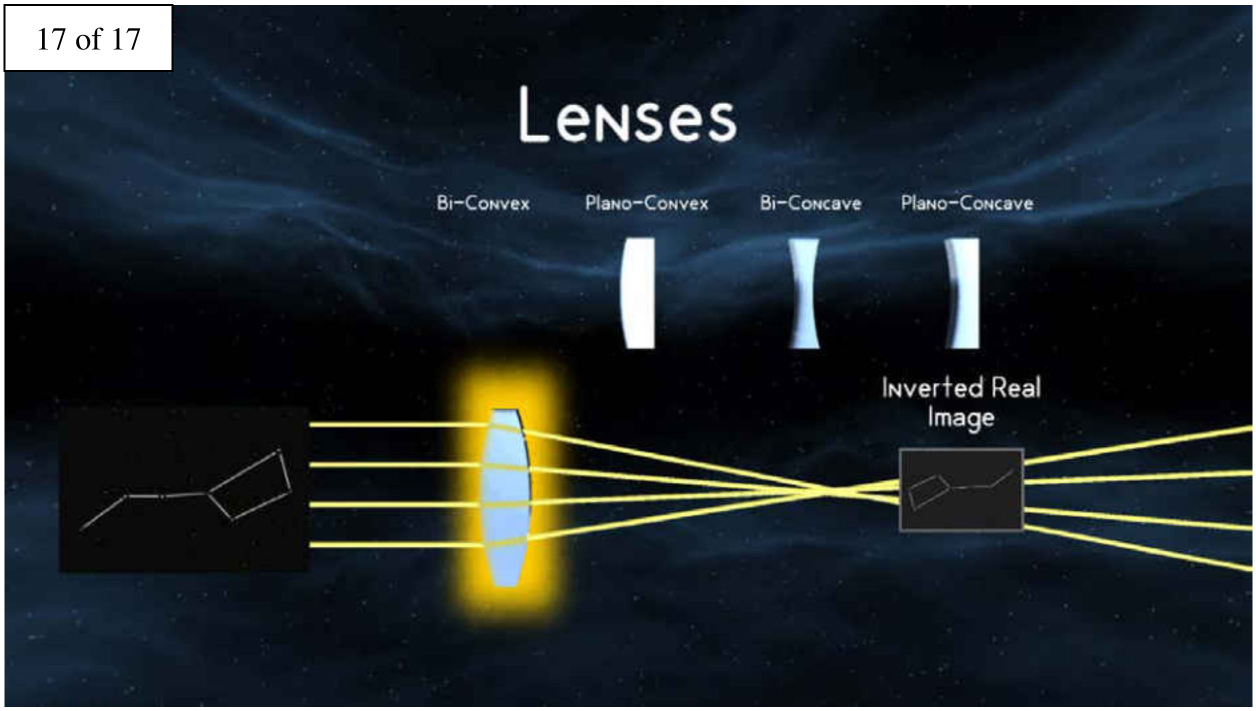


14 of 17




Lenses
Divergent (concave) lenses scatter light away and form a virtual image.
Bi-Concave
Plano-Concave


Lenses
Bi-Convex
Plano-Convex
Bi-Concave
Plano-Concave
Select the bi-convex lens


Lenses
Bi-Convex
Plano-Convex
Bi-Concave
Plano-Concave
Inverted Real
Image

APPENDIX D

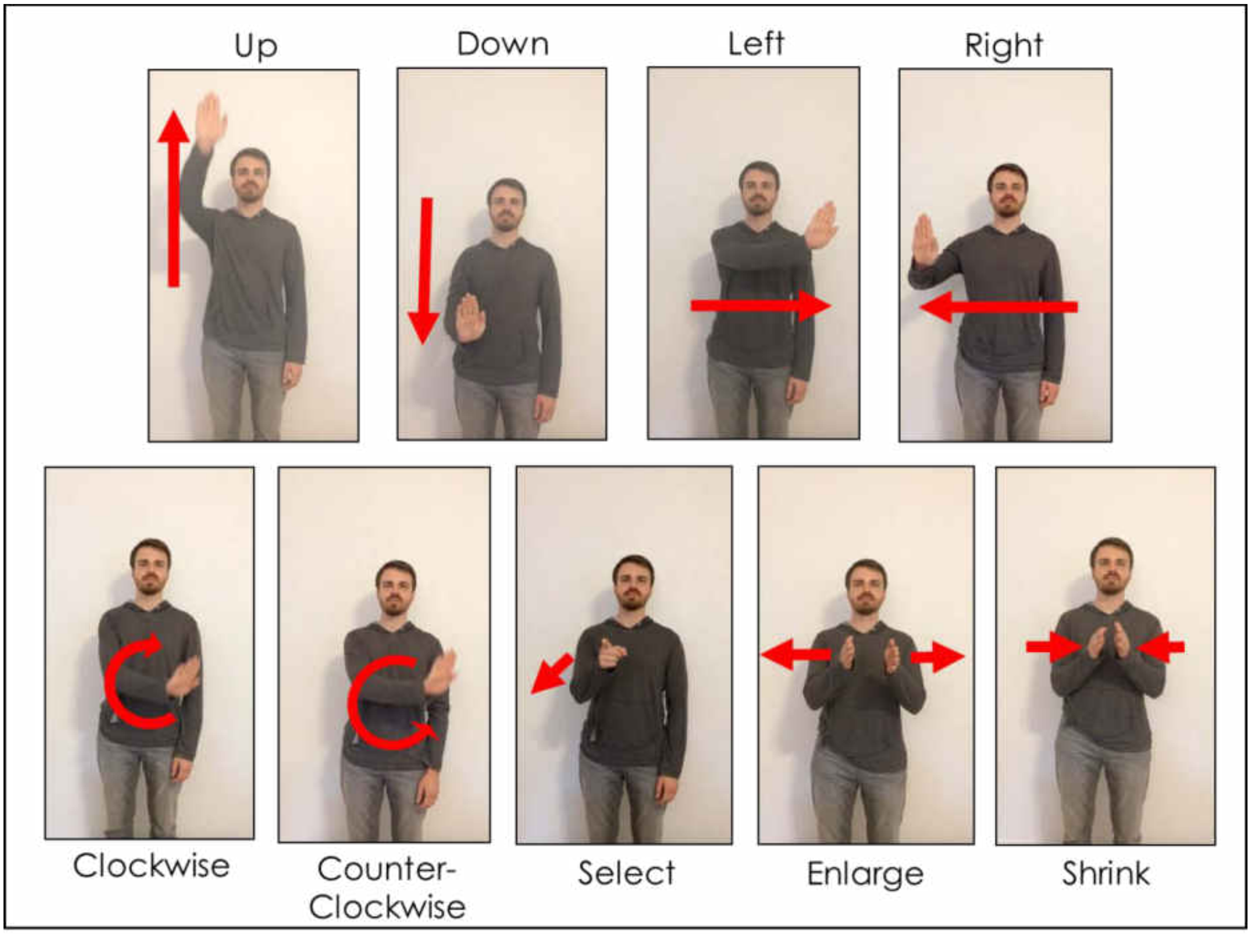
BRIEF ASSESSMENT OF GESTURE

Brief self-rating scale for the assessment of individual differences in gesture perception and production (BAG)

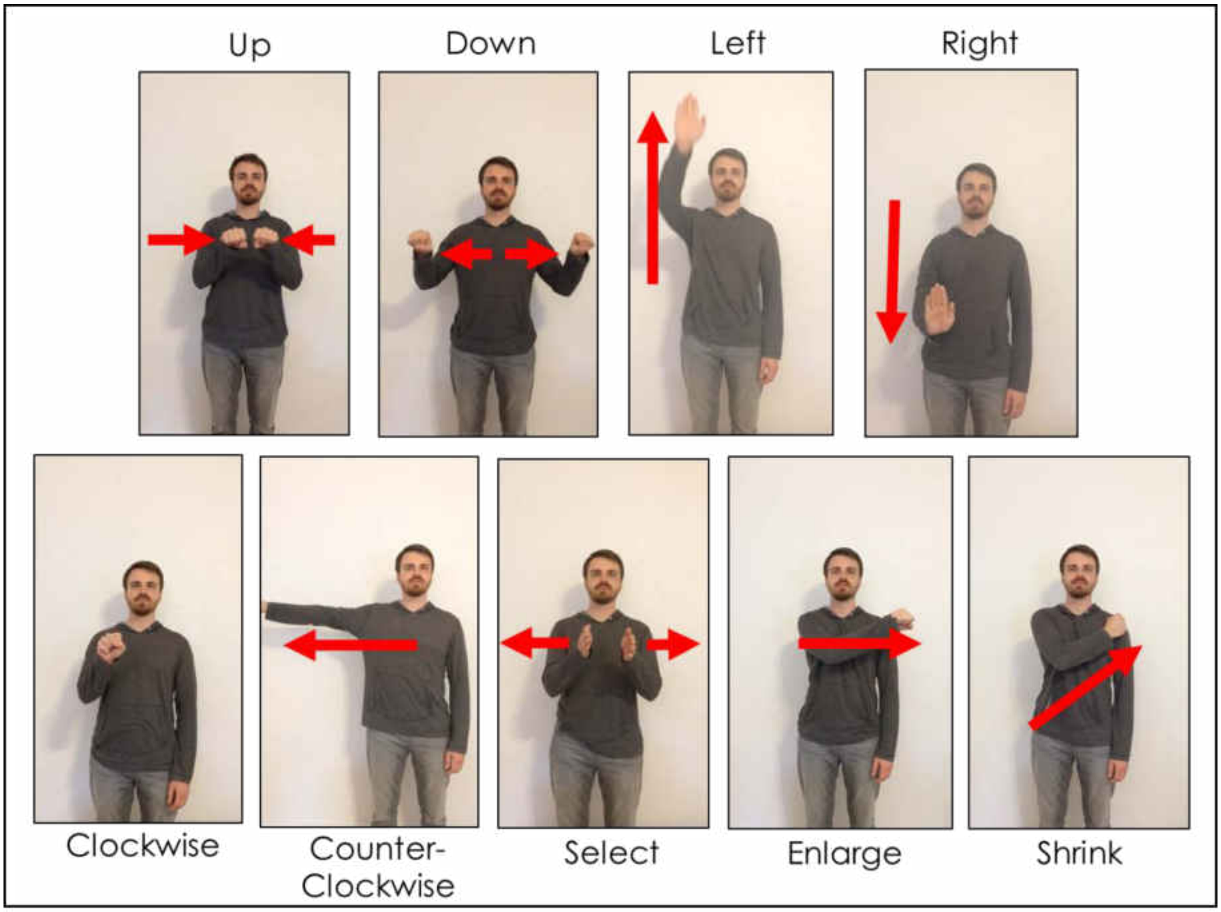
Item	not agree				fully agree
1. I usually gesture a lot when I talk to make myself understood better.	1	2	3	4	5
2. I like talking to people who gesture a lot when they talk.	1	2	3	4	5
3. I've been told before that I gesture a lot when I talk.	1	2	3	4	5
4. I find it more difficult to understand actors when they gesture a lot.	1	2	3	4	5
5. When talking in noisy places, I often gesture a lot to make myself understood over the noise.	1	2	3	4	5
6. I find it very annoying when I'm talking to someone who gestures a lot during the conversation.	1	2	3	4	5
7. I often feel amazed by people who are able to gesture a lot when they talk.	1	2	3	4	5
8. When I'm in a foreign country where I don't speak the language so well, I do a lot of gesturing to reinforce what I'm saying.	1	2	3	4	5
9. During a lecture, it's very distracting to me if the speaker gestures a lot.	1	2	3	4	5
10. When I see someone gesturing a lot, I often wonder if I would have used the same gestures.	1	2	3	4	5
11. It looks silly when you see a conversation between two people, and one of them is gesturing a lot.	1	2	3	4	5
12. I like having my hands free when I have a discussion with someone.	1	2	3	4	5

APPENDIX E

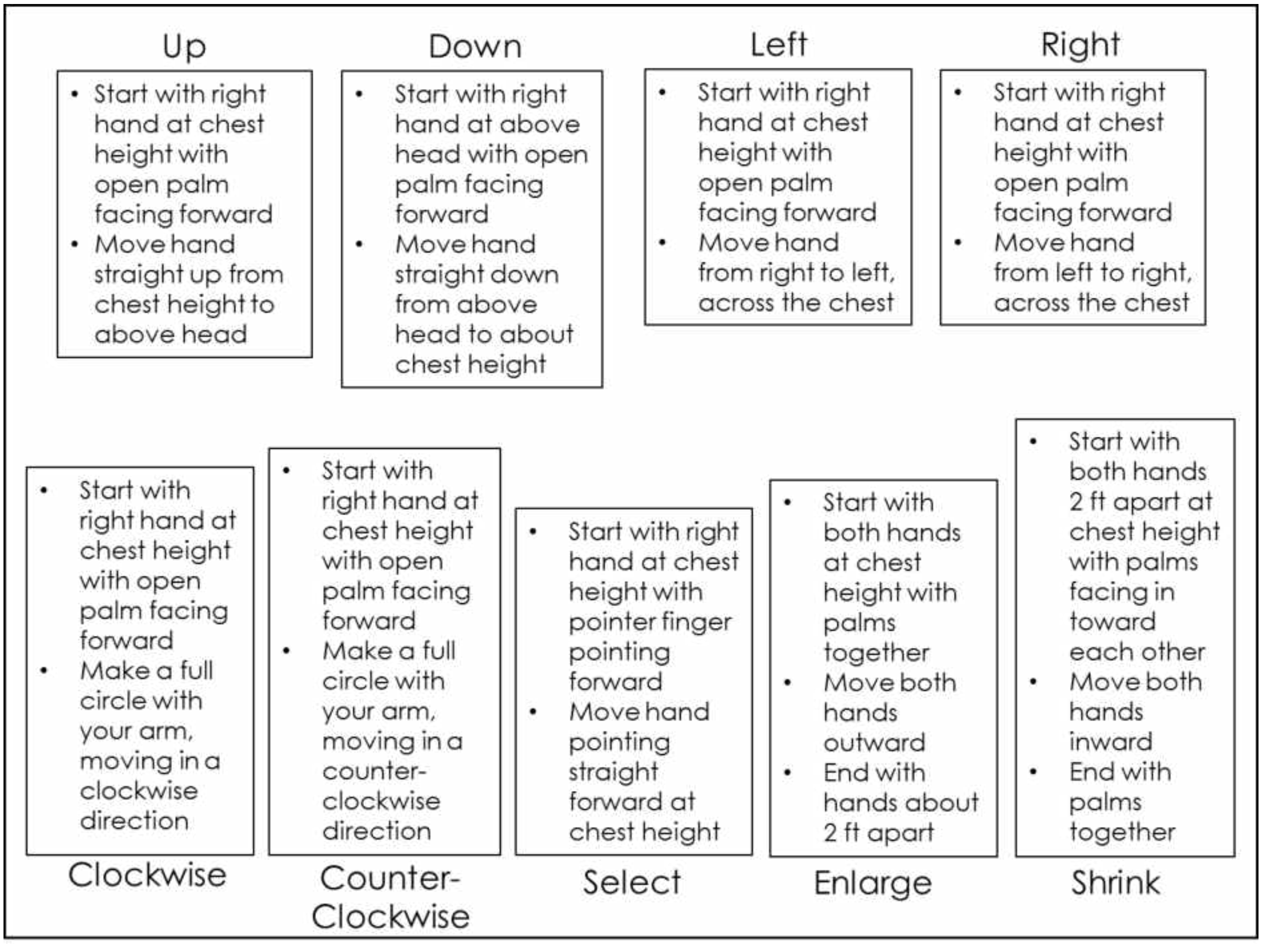
GESTURE REFERENCE SHEETS



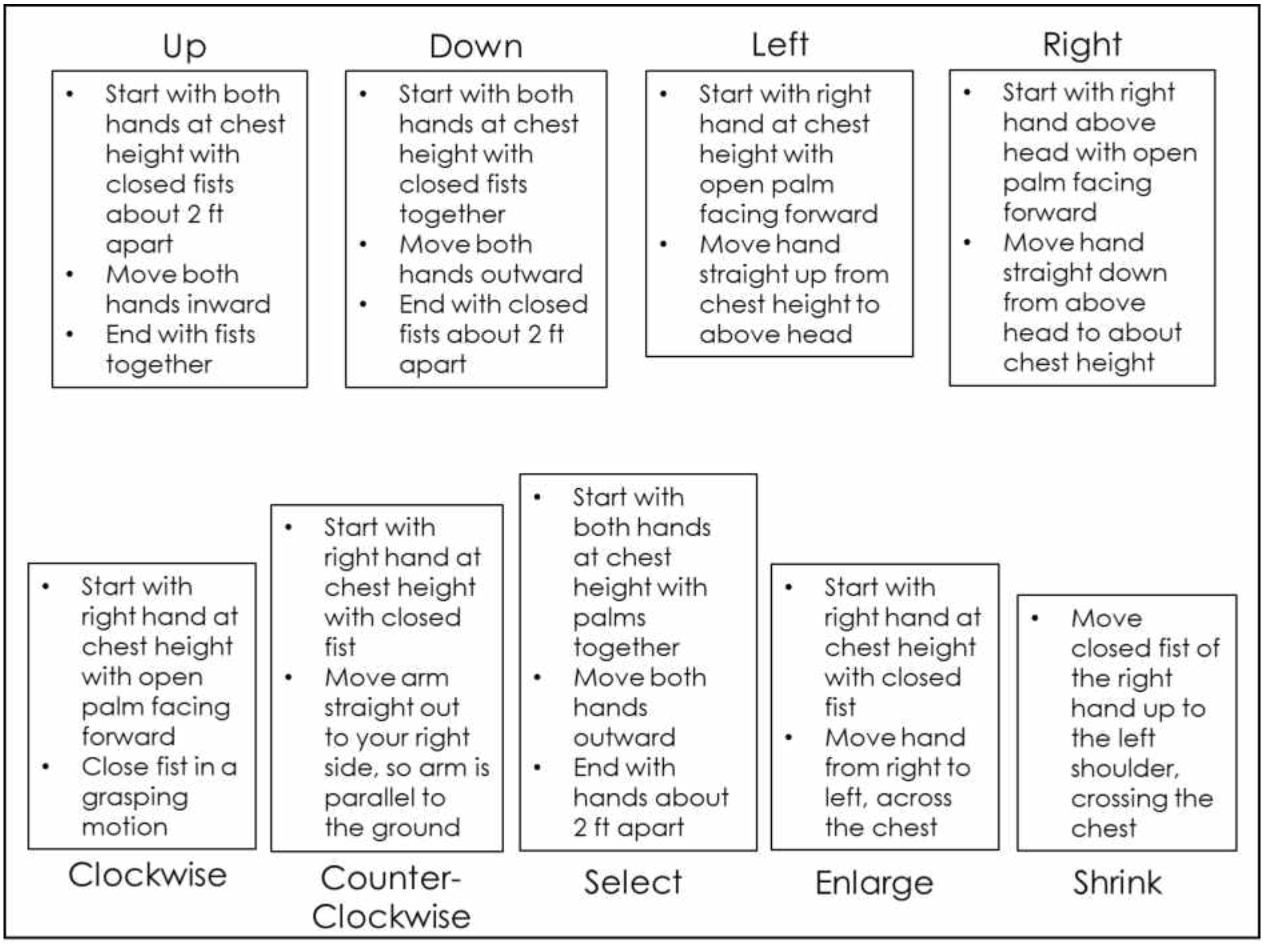
Gesture Reference Sheet with Natural Gestures for Video Instruction



Gesture Reference Sheet with Arbitrary Gestures for Video Instruction



Gesture Reference Sheet with Natural Gestures for Text Instruction



Gesture Reference Sheet with Arbitrary Gestures for Text Instruction

APPENDIX F

PRESENCE QUESTIONNAIRE

Presence Questionnaire

For the following questions, please indicate your response by circling the number that most appropriately describes your experience in the environment. Please consider the entire scale when making your responses, as the intermediate levels may apply. Answer the questions independently in the order that they appear. Do not skip questions or return to previous questions to change your answer.

1. How much were you able to control events?

1	2	3	4	5	6	7
Not at all			Moderately			Completely

2. How responsive was the environment to actions that you initiated (or performed)?

1	2	3	4	5	6	7
Not Responsive			Moderately Responsive			Completely Responsive

3. How natural did your interactions with the environment seem?

1	2	3	4	5	6	7
Extremely Artificial			Borderline			Completely Natural

4. How much did the visual aspects of the environment involve you?

1	2	3	4	5	6	7
Not at all			Somewhat			Completely

5. How much did the auditory aspects of the environment involve you?

1	2	3	4	5	6	7
Not at all			Somewhat			Completely

6. How natural was the mechanism which controlled movement through the environment?

1	2	3	4	5	6	7
Extremely Artificial			Borderline			Completely Natural

7. How compelling was your sense of objects moving through space?

1	2	3	4	5	6	7
Not at all			Moderately Compelling			Very Compelling

8. How much did your experiences in the virtual environment seem consistent with your real world experiences?

1	2	3	4	5	6	7
Not Consistent			Moderately Consistent			Very Consistent

9. How well could you identify sounds?

1	2	3	4	5	6	7
Not at all			Somewhat			Completely

10. How well could you localize sounds?

1	2	3	4	5	6	7
Not at all			Somewhat			Completely

11. How compelling was your sense of moving around inside the virtual environment?

1	2	3	4	5	6	7
Not Compelling			Moderately Compelling			Very Compelling

12. How much delay did you experience between your actions and expected outcomes?

1	2	3	4	5	6	7
No Delays			Moderate Delays			Long Delays

13. How proficient in moving and interacting with the virtual environment did you feel at the end of the experience?

1	2	3	4	5	6	7
Not Proficient			Reasonably Proficient			Very Proficient

14. How much did the visual display quality interfere or distract you from performing assigned tasks or required activities?

1	2	3	4	5	6	7
Not at all			Interfered Somewhat			Prevented Task Performance

15. How much did the control devices interfere with the performance of assigned tasks or with other activities?

1	2	3	4	5	6	7
Not at all			Interfered Somewhat			Interfered Greatly

16. How well could you concentrate on the assigned tasks or required activities rather than on the mechanisms used to perform those tasks or activities?

1	2	3	4	5	6	7
Not at all			Somewhat			Completely

17. How completely were your senses engaged in this experience?

1	2	3	4	5	6	7
Not at all			Somewhat			Completely

18. Were there moments during the virtual environment experience when you felt completely focused on the task or environment?

1	2	3	4	5	6	7
None			Few			Several

19. How easily did you adjust to the control devices used to interact with the virtual environment?

1	2	3	4	5	6	7
Not easily			Somewhat Easily			Very Easily

APPENDIX G

CORRELATION OF MEASURES

Table 20. Zero-order Correlations of Prescreening and Experimental Measures

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.
Prescreen Measures																
1. Pre-test	1															
2. PFT	.154	1														
3. Video Game	.045	.158	1													
4. BAG Perc	-.030	.023	.072	1												
5. BAG Prod	.095	.154	-.139	-.064	1											
6. BAG SocProd	.106	.198*	-.060	.094	.510**	1										
7. BAG SocPerc	-.061	-.012	.095	.140	.278**	.211*	1									
8. Time Pre-test	-.003	-.158	-.171	-.080	-.110	-.046	-.079	1								
Experiment Measures																
9. Delta	-.052	.352**	.182	.121	.217*	.179	.091	-.024	1							
10. ME	.019	-.139	-.064	-.096	-.035	-.139	-.088	-.072	-.045	1						
11. InstEff	.056	-.340**	-.171	-.151	-.179	-.225*	-.124	-.030	-.725**	.720**	1					
12. Time	-.280**	-.112	-.075	-.157	.112	-.059	-.049	-.072	.010	.429**	.282**	1				
13. SUS	.063	.041	-.102	.136	-.092	-.019	.108	-.061	.005	-.339**	-.231*	-.566**	1			
14. PQ Inv	.013	-.071	-.129	.004	-.006	-.031	.163	-.094	.044	-.047	-.066	-.236*	.598**	1		
15. PQ Sensory	.057	.018	-.114	-.026	.169	-.028	.158	.062	.139	.056	-.057	.026	.197*	.340**	1	
16. PQ AdaptImm	.149	.047	-.191	.046	-.176	-.115	.048	-.123	.010	-.104	-.082	-.281**	.581**	.601**	.345**	1
17. PQ Interface	-.088	-.249*	-.121	.054	.032	-.030	-.028	-.107	-.204*	.258**	.315**	.300**	-.414**	-.144	-.173	-.184

Note. $^{*}p$<.05, $^{**}p$<.01 (two-tailed); 1. Pretest = Knowledge of Optics Pre-test score; 2. PFT = Paper Folding Test; 3. Video Game = Video game hours per week; 4. BAG Perc = Brief Assessment of Gestures Perception; 5. BAG Prod = Brief Assessment of Gestures Production; 6. BAG SocProd = Brief Assessment of Gestures Social Production; 7. BAG Prod = Brief Assessment of Gestures Production; 8. Time Pre-test = Days from Pre-test; 9. Delta = Knowledge of Optics Delta score; 10. ME = Mental Effort; 11. InstEff = Instructional Efficiency; 12. Time = Time to complete lesson; 13. SUS = System Usability Scale; 14. PQ Inv = Presence Questionnaire Involvement; 15. PQ Sensory = Presence Questionnaire Sensory Fidelity; 16. PQ AdaptImm = Presence Questionnaire Adaptation/Immersion; 17. PQ Interface = Presence Questionnaire Interface Quality

APPENDIX H

IRB APPROVALS

IRB Approval: Study 1

University of Central Florida Institutional Review Board
Office of Research & Commercialization
12201 Research Parkway, Suite 501
Orlando, Florida 32826-3246
Telephone: 407-823-2901 or 407-882-2276
www.research.ucf.edu/compliance/irb.html

Approval of Human Research

From: **UCF Institutional Review Board #1**
FWA00000351, IRB00001138

To: **Shannon K. Bailey**

Date: **September 26, 2016**

Dear Researcher:

On 09/26/2016 the IRB approved the following human participant research until 09/25/2017 inclusive:

Type of Review:	UCF Initial Review Submission Form Expedited Review
Project Title:	Gesture-based Computer Interaction Pilot 1
Investigator:	Shannon K. Bailey
IRB Number:	SBE-16-12512
Funding Agency:	
Grant Title:	
Research ID:	N/A

The scientific merit of the research was considered during the IRB review. The Continuing Review Application must be submitted 30days prior to the expiration date for studies that were previously expedited, and 60 days prior to the expiration date for research that was previously reviewed at a convened meeting. Do not make changes to the study (i.e., protocol, methodology, consent form, personnel, site, etc.) before obtaining IRB approval. A Modification Form **cannot** be used to extend the approval period of a study. All forms may be completed and submitted online at https://iris.research.ucf.edu .

If continuing review approval is not granted before the expiration date of 09/25/2017, approval of this research expires on that date. When you have completed your research, please submit a Study Closure request in iRIS so that IRB records will be accurate.

Use of the approved, stamped consent document(s) is required. The new form supersedes all previous versions, which are now invalid for further use. Only approved investigators (or other approved key study personnel) may solicit consent for research participation. Participants or their representatives must receive a copy of the consent form(s).

All data, including signed consent forms if applicable, must be retained and secured per protocol for a minimum of five years (six if HIPAA applies) past the completion of this research. Any links to the identification of participants should be maintained and secured per protocol. Additional requirements may be imposed by your funding agency, your department, or other entities. Access to data is limited to authorized individuals listed as key study personnel.

In the conduct of this research, you are responsible to follow the requirements of the Investigator Manual.

On behalf of Sophia Dziegielewski, Ph.D., L.C.S.W., UCF IRB Chair, this letter is signed by:

Page 1 of 2

Kamille Chap

Signature applied by Kamille Chaparro on 09/26/2016 11:19:28 AM EDT

IRB Coordinator

IRB Approval: Study 2

University of Central Florida Institutional Review Board
Office of Research & Commercialization
12201 Research Parkway, Suite 501
Orlando, Florida 32826-3246
Telephone: 407-823-2901 or 407-882-2276
www.research.ucf.edu/compliance/irb.html

Approval of Exempt Human Research

From: **UCF Institutional Review Board #1**
FWA00000351, IRB00001138

To: **Shannon K. Bailey** and Co-PI: **Valerie K. Sims**

Date: **January 18, 2017**

Dear Researcher:

On 01/18/2017, the IRB approved the following activity as human participant research that is exempt from regulation:

Type of Review:	Exempt Determination
Project Title:	Gesture-based Computer Interaction: Pilot 2
Investigator:	Shannon K. Bailey
IRB Number:	SBE-16-12680
Funding Agency:	
Grant Title:	
Research ID:	N/A

This determination applies only to the activities described in the IRB submission and does not apply should any changes be made. If changes are made and there are questions about whether these changes affect the exempt status of the human research, please contact the IRB. When you have completed your research, please submit a Study Closure request in iRIS so that IRB records will be accurate.

In the conduct of this research, you are responsible to follow the requirements of the Investigator Manual.

On behalf of Sophia Dziegielewski, Ph.D., L.C.S.W., UCF IRB Chair, this letter is signed by:

Kamille Chap

Signature applied by Kamille Chaparro on 01/18/2017 07:06:54 PM EST

IRB Coordinator

Page 1 of 1

IRB Approval: Experiment

University of Central Florida Institutional Review Board
Office of Research & Commercialization
12201 Research Parkway, Suite 501
Orlando, Florida 32826-3246
Telephone: 407-823-2901 or 407-882-2276
www.research.ucf.edu/compliance/irb.html

Approval of Human Research

From: **UCF Institutional Review Board #1**
FWA00000351, IRB00001138

To: **Shannon K. Bailey**

Date: **June 27, 2017**

Dear Researcher:

On 06/27/2017 the IRB approved the following human participant research until 06/26/2018 inclusive:

Type of Review:	UCF Initial Review Submission Form Expedited Review
Project Title:	Gesture-based Computer Interaction: Experiment
Investigator:	Shannon K. Bailey
IRB Number:	SBE-17-13135
Funding Agency:	
Grant Title:	
Research ID:	N/A

The scientific merit of the research was considered during the IRB review. The Continuing Review Application must be submitted 30days prior to the expiration date for studies that were previously expedited, and 60 days prior to the expiration date for research that was previously reviewed at a convened meeting. Do not make changes to the study (i.e., protocol, methodology, consent form, personnel, site, etc.) before obtaining IRB approval. A Modification Form **cannot** be used to extend the approval period of a study. All forms may be completed and submitted online at https://iris.research.ucf.edu .

If continuing review approval is not granted before the expiration date of 06/26/2018, approval of this research expires on that date. When you have completed your research, please submit a Study Closure request in iRIS so that IRB records will be accurate.

Use of the approved, stamped consent document(s) is required. The new form supersedes all previous versions, which are now invalid for further use. Only approved investigators (or other approved key study personnel) may solicit consent for research participation. Participants or their representatives must receive a copy of the consent form(s).

All data, including signed consent forms if applicable, must be retained and secured per protocol for a minimum of five years (six if HIPAA applies) past the completion of this research. Any links to the identification of participants should be maintained and secured per protocol. Additional requirements may be imposed by your funding agency, your department, or other entities. Access to data is limited to authorized individuals listed as key study personnel.

In the conduct of this research, you are responsible to follow the requirements of the Investigator Manual.

On behalf of Sophia Dziegielewski, Ph.D., L.C.S.W., UCF IRB Chair, this letter is signed by:

Kanielle Chap

Page 1 of 2

Signature applied by Kamille Chaparro on 06/27/2017 09:20:05 AM EDT

IRB Coordinator

APPENDIX I

COPYRIGHT PERMISSIONS

<u>Brief Assessment of Gesture Permission</u>

License Number	4174951029785
License date	Aug 23, 2017
Licensed Content Publisher	Elsevier
Licensed Content Publication	Learning and Individual Differences
Licensed Content Title	A brief self-rating scale for the assessment of individual differences in gesture perception and production
Licensed Content Author	Arne Nagels,Tilo Kircher,Miriam Steines,Michael Grosvald,Benjamin Straube
Licensed Content Date	Apr 1, 2015
Licensed Content Volume	39
Licensed Content Issue	n/a
Licensed Content Pages	8
Type of Use	reuse in a thesis/dissertation
Portion	excerpt
Number of excerpts	1
Format	both print and electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No
Title of your thesis/dissertation	Getting the upper hand: Does intuitive gesturing result in better learning than embodied training on a conceptual task?
Expected completion date	Nov 2017
Estimated size (number of pages)	150
Requestor Location	Miss. Shannon Bailey 4000 Central Florida Blvd ORLANDO, FL 32816 United States Attn: Miss. Shannon Bailey
Publisher Tax ID	98-0397604
Total	0.00 USD

<u>Presence Questionnaire Permission</u>

Copyright Clearance Center

RightsLink®

Home | Account Info | Help

MIT Press Journals

Title:	The Factor Structure of the Presence Questionnaire
Author:	Bob G. Witmer, Christian J. Jerome, Michael J. Singer
Publication:	Presence: Teleoperators & Virtual Environments
Publisher:	MIT Press Journals
Date:	06/01/2005

Logged in as:
Shannon Bailey
Account #:
3001187698
LOGOUT

Thesis / Dissertation Use

Thesis / Dissertation reuse is permitted free of charge with the following restrictions:
* Reuse must be by an advanced degree candidate for use in his/her master's or doctoral thesis or dissertation.
* No more than 12 copies (or users) of thesis / dissertation containing MIT Press Journal content may be made.
* Future publications, for-profit publications, or inability to meet these restrictions require a for-fee license through this service.
* Full credit and proper copyright notice must be given for material used.

BACK | CLOSE WINDOW

REFERENCES

Abadi, H.G.N., Peng, L.Y., & Zadeh, A.M.H. (2012). Guessability study on considering cultural values in gesture design for different user interfaces. *In International Proceedings of Economics Development & Research.*

Alexander, K.P. (2013). The usability of print and online video instructions. *Technical Communications Quarterly, 22*, 237-259.

Alibali, M.W., Spencer, R.C., Knox, L., & Kita, S. (2011). Spontaneous gestures influence strategy choices in problem solving. *Psychological Science, 22*(9), 1138-1144.

Ayres, P. & Paas, F. (2007). Making instructional animations more effective: A cognitive load approach. *Applied Cognitive Psychology, 21,* 695-700.

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*(11), 417-423.

Baddeley, A.D., & Hitch, G. (1974). Working memory. *Psychology of Learning and Motivation, 8*, 47-89.

Barsalou, L.W. (2008). Grounded cognition. *Annual Review of Psychology, 59*, 617-645.

Barsalou, L.W., Barbey, A. K., Simmons, W., & Santos, A. (2005). Embodiment in religious knowledge. *Journal of Cognition and Culture, 5*(1-2), 14-57.

Betrancourt, M. (2005). The animation and interactivity principles in multimedia learning. In R. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp. 287-296). New York: Cambridge University Press.

Borghi, A.M. (2007). Object concepts and embodiment: Why sensorimotor and cognitive processes cannot be separated. *La nuova critica, 49*(50), 90-107.

Boyatzis, C.J., & Watson, M.W. (1993). Preschool children's symbolic representation of objects through gestures. *Child Development, 64*, 729-735.

Brooke, J. (1996). SUS-A quick and dirty usability scale. *Usability Evaluation in Industry, 189*, 4-7.

Bruner, J.S. (1966). On cognitive growth. In J.S. Bruner, P. Greenfield, & R. Olver (Eds.), *Studies in cognitive growth: A collaboration at the center for cognitive studies.* New York: Wiley.

Brünken, R., Seufert, T., & Paas, F. (2010). Measuring cognitive load. In J. L. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive load theory,* (29-47). New York: Cambridge University Press.

Capone, N.C. & McGregor, K.K. (2004). Gesture development: A review for clinical and research practices. *Journal of Speech, Language, and Hearing Research, 47*, 173-186.

Cook, S.W., Mitchell, Z., & Goldin-Meadow, S. (2008). Gesturing makes learning last. *Cognition, 106*(2), 1047-1058.

Davidson, M.W. (2015a, November 13). *Introduction to Lenses.* Retrieved from http://micro.magnet.fsu.edu/optics/lightandcolor/lenses.html

Davidson, M.W. (2015b, November 13). *Reflection of Light.* Retrieved from http://micro.magnet.fsu.edu/optics/lightandcolor/reflection.html

Davidson, M.W. (2015c, November 13). *Refraction of Light.* Retrieved from http://micro.magnet.fsu.edu/optics/lightandcolor/refraction.html

Dodds, T.J., Mohler, B.J., & Bulthoff, H.H. (2011). Talk to the virtual hands: Self-animated avatars improve communications in head-mounted display virtual environments. *PLoS ONE, 6*(10), e25759.

Ekstrom, R., French, J., Harman, H., & Dermen, D. (1976). *Manual for kit of factor referenced cognitive tests*. Princeton, NJ: Educational Testing Service.

Engelkamp, J., & Jahn, P. (2003). Lexical, conceptual and motor information in memory for action phrases: A multi-system account. *Acta Psychologica, 113*, 147-165.

Engelkamp, J., & Zimmer, H.D. (1997). Sensory factors in memory for subject-performed tasks. *Acta Psychologica, 96*, 43-60.

Faul, F., Erdfelder, E., Lang, A.G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175-191.

Field, A. (2013). Discovering statistics using IBM SPSS statistics. Sage.

Garber, P. & Goldin-Meadow, S. (2002). Gesture offers insight into problem-solving in adults and children. *Cognitive Science, 26*, 817-831.

Garbarini, F., & Adenzato, M. (2004). At the root of embodied cognition: Cognitive science meets neurophysiology. *Brain and Cognition, 56*(1), 100-106.

Ginns, P. (2005). Meta-analysis of the modality effect. *Learning and Instruction, 15*, 313-331.

Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., & Wagner, S. (2001). Explaining math: Gesturing lightens the load. *Psychological Science, 12*(6), 516-522.

Goodale, M.A., & Humphrey, K.G. (1998). The objects of action and perception. *Cognition, 67*(1), 181-207.

Grandhi, S.A., Joue, G., & Mittelberg, I. (2010). Understanding naturalness and intuitiveness in gesture production: Insights for touchless gestural interfaces. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 821-824). ACM.

Hamblin, C.J. (2005). *Transfer of training from virtual reality environments* (Doctoral dissertation, Wichita State University).

Hegarty, M., Mayer, S., Kriz, S., & Keehner, M. (2005). The role of gestures in mental animation. *Spatial Cognition and Computation, 5*(4), 333-356.

Hegarty, M., & Sims, V. K. (1994). Individual differences in mental animation during mechanical reasoning. *Memory & Cognition, 22*(4), 411-430.

Henderson, T. (1999a). *Reflection and the Ray Model of Light.* Retrieved from http://www.physicsclassroom.com/class/refln

Henderson, T. (1999b). *Refraction and the Ray Model of Light.* Retrieved from http://www.physicsclassroom.com/class/refrn

Höffler, T.N. (2010). Spatial ability: Its influence on learning with visualizations – a meta-analytic review. *Educational Psychology Review, 22*, 245-269.

Hostetter, A.B. & Alibali, M.W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin and Review, 15*(3), 495-514.

Johnson-Glenberg, M.C., Birchfield, D.A., Tolentino, L., & Koziupa, T. (2014). Collaborative embodied learning in mixed reality motion-capture environments: Two science studies. *Journal of Educational Psychology, 106*(1), 86-104.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist, 38*(1), 23-31.

Kohler, E., Keysers, C., Umilta, M. A., Fogassi, L., Gallese, V., & Rizzolatti, G. (2002). Hearing sounds, understanding actions: action representation in mirror neurons. *Science, 297* (5582), 846-848.

Leppink, J., Paas, F., van Gog, T., van der Vleuten, C.P., & Merriënboer, J.J. (2014). Effects of pairs of problems and examples on task performance and different types of cognitive load. *Learning and Instruction, 30*, 32-42.

Leppink, J., Paas, F., Van der Vleuten, C. P., Van Gog, T., & Van Merriënboer, J. J. (2013). Development of an instrument for measuring different types of cognitive load. *Behavior Research Methods, 45*(4), 1058-1072.

Leppink, J., & van den Heuvel, A. (2015). The evolution of Cognitive Load Theory and its application to medical education. *Perspectives on Medical Education, 4*(3), 119-127.

Longo, M.R., & Lourenco, S.F. (2007). Space perception and body morphology: Extent of near space scales with arm length. *Experimental Brain Research, 177*(2), 285-290.

Mauney, D., Howarth, J., Wirtanen, A., & Capra, M. (2010). Cultural similarities and differences in user-defined gestures for touchscreen user interfaces. *In CHi'10 extended abstracts on Human Factors in Computing Systems* (pp. 4015-4020). ACM.

Mayer, R.E. & Anderson, R.B. (1991). Animations need narrations: An experimental test of the dual-coding hypothesis. *Journal of Educational Psychology, 83*(4), 484-490.

Mayer, R.E. & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*(1), 43-52.

Mayer, R.E. & Moreno, R. (2010). Techniques that reduce extraneous cognitive load and manage intrinsic cognitive load during multimedia learning. In J. L. Plass, R. Moreno, &

R. Brünken (Eds.), *Cognitive load theory,* (29-47). New York: Cambridge University Press.

Moreno, R. & Park, B. (2010). Cognitive Load Theory: Historical development and relation to other theories. In J. L. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive load theory,* (29-47). New York: Cambridge University Press.

Mayer, R.E., & Sims, V.K. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology, 86*(3), 389.

Nagels, A., Kircher, T., Steines, M., Grosvald, M., & Straube, B. (2015). A brief self-rating scale for the assessment of individual differences in gesture perception and production. *Learning and Individual Differences, 39*, 73-80.

Nielsen, M., Störring, M., Moeslund, T.B., & Granum, E. (2003, April). A procedure for developing intuitive and ergonomic gesture interfaces for HCI. *In International Gesture Workshop* (pp. 409-420). Springer Berlin Heidelberg.

Nilsson, L.G., Cohen, R.L., & Nyberg, L. (1989). Recall of enacted and nonenacted instructions compared: Forgetting functions. *Psychological Research, 51*(4), 188-193.

Norman, D.A. (2002). *The design of everyday things.* New York: Basic Books.

Paas, F., & van Merriënboer, J.J.G. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of Educational Psychology, 86*(1), 122-133.

Paas, F., van Merriënboer, J.J. G., & Adam, J.J. (1994). Measurement of cognitive load in instructional research. *Perceptual and Motor Skills, 79*, 419–430.

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. (2003). Cognitive load measurement as a means to advance Cognitive Load Theory. *Educational Psychologist, 38*(1), 63-71.

Pezzulo, G., Barsalou, L.W., Cangelosi, A., Fischer, M.A., McRae, K., & Spivey, M. (2011). The mechanics of embodiment: A dialogue on embodiment and computational modeling. *Frontiers in Psychology, 2*(5), 1-21.

Procci, K., James, N., & Bowers, C. (2013, September). The effects of gender, age, and experience on game engagement. In *Proceedings of the Human Factors and Ergonomics Society annual meeting* (Vol. 57, No. 1, pp. 2132-2136). Los Angeles, CA: SAGE Publications.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience, 27*, 169-192.

Roupé, M., Bosch-Sijtsema, P., & Johansson, M. (2014). Interactive navigation interface for virtual reality using the human body. *Computers, Environment and Urban Systems, 43*, 42-50.

Schroeder, B.L., Bailey, S.K., Johnson, C.I., & Gonzalez-Holland, E. (2017, July). Presence and usability do not directly predict procedural recall in virtual reality training. In *International Conference on Human-Computer Interaction* (pp. 54-61). Springer.

Schürmann, T., Binder, C., Janzarik, G., & Vogt, J. (2015). Movement transformation on multi-touch devices: Intuition or instructional preparation? *Applied Ergonomics, 50*, 251-255.

Schwartz, R.N., & Plass, J.L. (2014). Click versus drag: User-performed tasks and the enactment effect in an interactive multimedia environment. *Computers in Human Behavior, 33*, 242-255.

Science Voyages Level Green: Exploring the Life, Earth, and Physical Sciences (illustrated ed.). (1999). New York: Glencoe/ McGraw-Hill.

Sheu, F.R., & Chen, N.S. (2014). Taking a signal: A review of gesture-based computing research in education. *Computers and Education, 78*, 268-277.

Shiratuddin, M.F., & Wong, K.W. (2011). Non-contact multi-hand gestures interaction techniques for architectural design in a virtual environment. *Proceedings of the 2011 International Conference on Information Technology and Multimedia*.

Silpasuwanchai, C., & Ren, X. (2015). Designing concurrent full-body gestures for intense gameplay. *International Journal of Human-Computer Studies, 80*, 1-13.

Singer, A.S., & Goldin-Meadow, S. (2005). Children learn when their teacher's gestures and speech differ. *Psychological Science, 16*(2), 85-89.

Sousa-Poza, J.F., Rohrberg, R., & Mercure, A. (1979). Effects of type of information (abstract-concrete) and field dependence on asymmetry of hand movements during speech. *Perceptual and Motor Skills, 48*, 1323-1330.

Steuer, J.S. (1992). Defining virtual reality: Dimensions determining telepresence. *Journal of Communication 42*(4), 73–93.

Sweller, J. (2010). Cognitive load theory: Recent theoretical advances. In J. L. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive load theory,* (29-47). New York: Cambridge University Press.

Sweller, J., van Merriënboer, J.J.G., & Paas, F.G.W.C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251–296.

Tabachnick, B.G., & Fidell, L.S. (2013). *Using Multivariate Statistics* (5th edn). Boston: Pearson Education.

Terlecki, M.S., & Newcombe, N.S. (2005). How important is the digital divide? The relation of computer and videogame usage to gender differences in mental rotation ability. *Sex Roles, 53*(5), 433-441.

Ullrich, D., & Diefenbach, S. (2010). From magical experience to effortlessness: an exploration of the components of intuitive interaction. *In Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*, 801-804.

Vallotton, C., Fusaro, M., Hayden, J., Decker, K., & Gutowski, E. (2015). Give me a hand: Differential effects of gesture type in guiding young children's problem-solving. *Instructional Science, 43*(6), 709-735.

Van Dam, A. (1997). Post-WIMP user interfaces. *Communications of the ACM, 40*(2), 63-67.

Van Gog, T., Paas, F., Marcus, N., Ayres, P., & Sweller, J. (2009). The mirror-neuron system and observational learning: Implications for the effectiveness of dynamic visualizations, *Educational Psychology Review, 21,* 21-30.

Wagner, S.M., Nusbaum, H., & Goldin-Meadow, S. (2004). Probing the mental representation of gesture: Is handwaving spatial? *Journal of Memory and Language, 50*(4), 395-407.

Weidinger, N., Lindner, K., Hogrefe, K., Ziegler, W., & Goldenberg, G. (2017). Getting a grasp on children's representational capacities in pantomime of object use. *Journal of Cognition and Development, 18*(2), 246-269.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin and Review, 9*(4), 625-636.

Witmer, B.G., Jerome, C.J., & Singer, M.J. (2005). The factor structure of the Presence Questionnaire. *Presence: Teleoperators and Virtual Environments, 14*(3), 298-312.

Zacks, J.M, & Tversky, B. (2003). Structuring information interfaces for procedural learning. *Journal of Experimental Psychology: Applied, 9*(2), 88-100.