Exploring the Efficacy of Using Augmented Reality to Alleviate Common Misconceptions about Natural Selection

by

Matthew "Hue" Henry

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved July 2019 by the
Graduate Supervisory Committee:

Robert K. Atkinson, Chair
Mina Johnson-Glenberg
Brian C. Nelson

ARIZONA STATE UNIVERSITY

August 2019

ABSTRACT

Evidence suggests that Augmented Reality (AR) may be a powerful tool for alleviating certain, lightly held scientific misconceptions. However, many misconceptions surrounding the theory of evolution are deeply held and resistant to change. This study examines whether AR can serve as an effective tool for alleviating these misconceptions by comparing the change in the number of misconceptions expressed by users of a tablet-based version of a well-established classroom simulation to the change in the number of misconceptions expressed by users of AR versions of the simulation.

The use of realistic representations of objects is common for many AR developers. However, this contradicts well-tested practices of multimedia design that argue against the addition of unnecessary elements. This study also compared the use of representational visualizations in AR, in this case, models of ladybug beetles, to symbolic representations, in this case, colored circles.

To address both research questions, a one-factor, between-subjects experiment was conducted with 189 participants randomly assigned to one of three conditions: non-AR, symbolic AR, and representational AR. Measures of change in the number and types of misconceptions expressed, motivation, and time on task were examined using a pair of planned orthogonal contrasts designed to test the study's two research questions.

Participants in the AR-based condition showed a significantly smaller change in the number of total misconceptions expressed after the treatment as well as in the number of misconceptions related to intentionality; none of the other misconceptions examined showed a significant difference. No significant differences were found in the total

number of misconceptions expressed between participants in the representative and symbolic AR-based conditions, or on motivation. Contrary to the expectation that the simulation would alleviate misconceptions, the average change in the number of misconceptions expressed by participants increased. This is theorized to be due to the juxtaposition of virtual and real-world entities resulting in a reduction in assumed intentionality.

DEDICATION

This work is dedicated to my amazing wife, Lainie, for the love and support she provided throughout this process. She is the love of my life and the inspiration for all that I do.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my family, especially my parents and my grandfather, for always encouraging me to seek out the answers to whatever questions I may have. Thank you for putting up with all of the “what if” questions I asked as a child, and for showing me how to do the research and experimentation needed to find the answers on my own. Thank you for making me who I am.

I also wish to thank my committee for the inspiration and guidance they have provided. From the moment I started on this journey, one person has always been there for me and gone out of her way to provide me with the tools I needed to succeed. Even before I enrolled in the program, Dr. Mina C. Johnson-Glenberg opened up opportunities for me that I would not have otherwise. I cannot thank her enough for her guidance and assistance. I would not be at this point if not for her. Likewise, thank you to Dr. Robert K. Atkinson for taking on the role of my committee chair and helping me to navigate the paperwork and deadlines needed to complete this process, and to Dr. Brian Nelson for providing valuable input and advice on how to best approach this study.

Finally, thank you to everyone who encouraged, inspired, and supported me along the way. Thank you.

TABLE OF CONTENTS

Page

Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

Chapter 1

Introduction

Augmented Reality

Augmented Reality, often abbreviated as AR, is commonly defined as any technology which presents virtual objects over the real-world, thus causing virtual objects to appear to exist in the same space as objects in the real world (Azuma, 1997). AR is a rapidly advancing technology that is gaining popularity in both the commercial and educational markets (Bower et al., 2014).

The IT research and advisory firm Gartner, Inc. publishes an annual report and corresponding graphic designed to represent "the maturity and adoption of technologies." (https://www.gartner.com/technology/research/methodologies/hype-cycle.jsp). According to the 2017 report, AR is quickly moving beyond the "trough of disillusionment" where a technology will either "shake out or fail," and is entering the "plateau of productivity", a state where mainstream adoption takes place and the technology becomes commonplace, within the next five to ten years (Panetta, 2017).

Similarly, over the past decade, the number of studies on the use of AR for education is growing rapidly. The use of AR technology has been studied for a variety of topics, including science, mathematics, language learning, and visual art appreciation. (Chen et al., 2017). These studies have found AR to be effective in areas where students are expected to learn information that cannot be seen in the real world or without a specialized device, and in areas where students are expected to learn abstract or complex concepts. AR technology allows learners to observe and interact with otherwise invisible mechanisms, while providing structures that focus the learner’s attention on the relevant information, including dynamic changes in the phenomenon over time and scientific details that might otherwise be overlooked or unavailable (Yoon & Wang, 2014). AR technologies have likewise been identified as a key emerging technology for elementary and secondary education over the next five years (Johnson, Levine, Smith, & Haywood, 2010). It is critical that the efficacy and best practices for putting this new technology to use are thoroughly examined before AR becomes a mainstream tool for education.

Augmented Reality and Science Misconceptions

One area in which AR shows promise is in challenging and redressing of common science misconceptions (Yoon, Anderson, Lin & Elinich, 2017). Yoon et al. (2017) found that AR can help enable conceptual understanding of challenging scientific content and alleviate common misconceptions surrounding physics topics, such as Bernoulli's Principle. In one experiment, Yoon et al. (2017) provided museumgoers with an AR experience that allowed the participants to see, in real time, a visual representation of the air speed and pressure of two currents of air. This allowed participants to directly observe how the inverse relationship between the speed of the air and the air pressure in the room allows a real-world ball to remain floating in a stream of fast-moving air. The common misconception challenged by this AR simulation was the assumption that the relationship between air speed and air pressure is a direct relationship, when in fact, an inverse relationship is present; as air speed increases, air pressure decreases.

Although designed to study the use of AR in an informal learning environment (a science museum) rather than a traditional classroom, pretests and posttests confirmed that participants in the AR-condition scored in higher levels of understanding, and post-intervention interviews showed that participants were able to reason accurately about inverse relationships between air speed and pressure; their misconceptions about the relationship had been alleviated. Yoon et al. (2017) theorized that AR technology was effective at relieving this misconception due to AR's ability to provide an interactive environment in which normally imperceptible mechanisms can be made perceptible, and thus provide the scaffolding required to challenge the learner's intuitive misconceptions about the relationship.

However, many scientific misconceptions are not as simple as the misassumption of a direct versus inverse relationship. For example, in the study of Natural Selection, there are multiple common misconceptions that act as cognitive barriers to prevent naive learners from forming an accurate understanding of evolution and natural selection. These misconceptions include a bias towards a teleological or "purpose-seeking" view of the topic, an assumption that the entities involved in the process act with intention, the belief that attributes altered over the lifetime of an individual are always passed on to the next generation, and the categorization of evolution as a complex event, and not as a process or equilibration. These misconceptions are not mutually-exclusive and often highly correlated (Gregory, 2009; Ferrari & Chi, 1998).

Although these misconceptions differ from a simple inverted mathematical relationship, it is possible that AR can still help to challenge and alleviate these misconceptions in precisely the same way – by providing an interactive environment in which normally imperceptible mechanisms can be made perceptible, without the extraneous cognitive load needed for students to transfer their attention from the real-world to the virtual, as would be required in a traditional non-augmented environment.

The Importance of Evolution in Science Education

The difficulties of teaching the theory of evolution, are not a new field of study. A 1998 National Academy of Science report, *Teaching about Evolution and the Nature of Science* (1998, National Academy Press, Washington, DC), argued that the teaching of evolution is essential to school curricula if students are to understand biology. In a summary of this NAS report Meagher (1999), states that understanding of evolution is central not only to general biological understanding, but also critical for understanding concepts in molecular biology, developmental biology, physiology and anatomy, neurobiology and behavior, even into applications outside of biology including medicine and computer and systems applications (Meagher, 1999). Despite this long-standing support for the theory, misconceptions about the nature of evolution and natural selection are common, not only amongst naive learners, but even among graduate level biology students (Gregory & Ellis, 2009).

Common Misconceptions of Natural Selection

In the paper *Understanding Natural Selection: Essential Concepts and Common Misconceptions,* Gregory (2009) catalogs a list of common misconceptions associated with the theory of evolution and natural selection. These misconceptions are non-mutually exclusive and often correlated (Gregory, 2009; Ferrari & Chi, 1998). For the purposes of this study, four misconceptions were chosen due to their relevance to the intervention. The following common misconceptions were the focus of the intervention:

1. *Teleology*: One important element of evolutionary theory is that evolution is a two-step process; new traits arise due to random mutations but remain due to non-random forces of natural selection. However, it is not uncommon for naive learners of the theory to express only the non-random step of this two-step process. One common expression of this misconception is that learners come to believe that evolutionary forces occur in response to a particular need or predetermined plan, and that these mutations are always beneficial. However, in a correct interpretation of evolutionary theory, new traits arise in an undirected fashion; some are beneficial, some are neutral, and some are even detrimental to the survival of the species, and it is possible for even the neutral or detrimental traits to carry on to future generations. This misconception is the source of statements such as "cheetahs evolved so they can catch gazelles," or "finches diversified so they could eat different foods," or as "new traits always benefit the species."

2. *Anthropomorphism / Intentionality*: Another common misconception is the assigning of human-like conscious intent to the objects of natural selection or to the process itself. This misconception is the source of false statements such as "bacteria choose to become resistant to antibiotics" or "female gazelles choose the fastest males to produce faster offspring." This misconception can also be phrased as "natural selection involves a will, effort, or intent on the part of the organism/species."

3. *Soft Inheritance / Lamarckian Evolution*:

   A third misconception, often labeled after biologist Jean-Baptiste Lamarck, is that changes to an individual within its lifetime are passed on to offspring. A variation on this misconception is the idea that traits are gained through use or disuse – used traits remain, and those that are not used are lost. This misconception is the source of false statements such as "insects that changed colors to better match their surroundings survived, and so the species evolved to be darker colors" or "giraffes who stretched their necks to reach higher leaves were better fed than those who did not, and thus passed these stretched necks on to their offspring." This misconception can also be phrased as "Acquired characteristics can be inherited" or "Individual organisms can evolve during a single lifespan."

4. *Event-based (versus Equilibration-based) Ontology*: Ferrari and Chi (1998) have theorized that the source of many of these misconceptions is not the inability of students to comprehend individual principles, but instead, a general miscategorization of evolution as a complex event, and not as a process or equilibration. Chi (1997) proposes that physical processes, such as the process of natural selection, can be categorized into two types: events and equilibrations. Events are distinct, sequential actions with a beginning and end, are causal in nature, goal oriented, and contingent on other events that act as causes. Equilibrations, on the other hand, are uniform, simultaneous actions that are unbounded, perpetually ongoing, and independent of other processes

(Ferrari & Chi, 1998). For example, gravity is an equilibration – it is always occurring, affects all matter simultaneously and with the same, uniform force, and is independent of other processes. It would be a miscategorization and misconception to say that "gravity started to pull on an object because of a specific goal or cause." Likewise, it would be a miscategorization to refer to evolutionary forces as an event with a start and end point. The misconception that evolution is an event, and not the correct understanding of evolution as an equilibration, leads to false statements as "the moths began to evolve," or "humans are no longer evolving." This misconception can also be phrased as "Evolution is an event that have a start point, caused by other events, and end when they reach a goal."

A Constructivist Approach to Alleviating Misconceptions

The constructivist approach to education is the view that students construct their knowledge from individual experiences and from reasoning about those experiences (Hewson & Hewson, 2003). One of the barriers many students face when learning about evolution is that their previous experiences and reasoning about those experiences has led to prior conceptions that conflict with the theory. Thus, students of natural selection must do more than just add to their existing knowledge; learners must also revise their mental models of the world and create a new way of seeing. This type of learning is referred to as conceptual change (Sinatra, Brem, & Evans, 2008). The literature has shown that computer simulations which use a constructivist approach and meet four essential conditions can be powerful tools for enabling this type of conceptual change (Hewson & Hewson, 2003; Perkins & Simmons, 1988).

These four essential conditions are:

1. *Dissatisfaction*: Students must experience a sufficient level of dissatisfaction with their existing conception;
2. *Intelligible*: Students must find the new conception to be intelligible, and sufficiently easy to understand;
3. *Plausible*: Students must find the new conception to be plausible, based on their own personal experiences and knowledge; and
4. *Fruitful*: Students must find the new conception to be fruitful. They must be able to see how the new conception can be used to solve problems or predict phenomena.

Simulations, either computer based or otherwise, that meet these conditions, have been found to be useful in alleviating misconceptions (Windschitl & Andre, 1998). A popular non-computer simulation, designed by Stebbins and Allen (1975) for use in a biology classroom, meets these conditions, and was the inspiration and foundation for this study.

A Well-Established Simulation of Natural Selection

The challenges of overcoming misconceptions surrounding the topic of natural selection are not new; numerous lesson plans and classroom activities have been developed over the last few decades. One of these activities, developed by Stebbins and Allen (1975), uses physical manipulatives to simulate the process of natural selection.

Five hundred paper chips, 50 of each of 10 different colors, are created using colored construction paper and a quarter-inch paper punch. These colored chips simulate "prey" creatures in the simulation. A large, multi-colored fabric, which represents a "habitat" for the prey chips, is spread across a desk or table. This habitat fabric is intended to contain patterns that simulate natural environments, such as floral, leaf, or fruit prints, and contain a predominant color tone, against which one or more of the paper colors blend in. Ten chips of each color, one hundred chips total, are placed on the habitat image.

At the teacher’s instruction, one or more students, acting as "predators", are instructed to pick up one chip at a time and place it in a nearby bowl. Predators are assigned a quota of chips to capture, such that exactly 25% of the population remains once all quotas are met, and they are required to use only vision to locate chips to capture.

The surviving 25 chips are removed from the fabric habitat and grouped according to color. The number of survivors of each color are recorded. For each surviving chip, three chips of the same color are added to the pool, returning the population total to 100. These 100 chips are mixed and redistributed on the fabric and the process repeats. The process can be repeated as many times as the instructor feels necessary, and, generally, results in an obvious and observable shift in the color distribution of each generation. It is not uncommon for the majority of the population to closely match the habitat in fewer than three generations. (Stebbins & Allen, 1975).

Although the original description of this simulation dates back to 1975, variations of this lesson are still being used in biology classrooms today. One variation of the simulation using the metaphor of an imaginary organism was proposed by Fifield and Fall (1992). This version of the simulation was extended to include the concepts of sexual reproduction and Mendelian inheritance. In addition to using multiple colors, this version of the simulation also varies the chips by size and each chip is printed with symbols representing the genetics associated with the fictional organism's size and color. These additions are meant to help learners understand the difference between dominant and recessive genes. As with the Stebbins and Allen simulation, they did not measure the efficacy of the simulation, only its ability to accurately represent the scientific concept of natural selection (Fifield & Fall, 1992).

Geraedts and Boersma (2006) also used the simulation as one element in a series of lessons designed to eliminate the common misconception that individuals can pass on acquired characteristics to their offspring. These lessons consisted of three parts. First, learners answered a series of questions designed to test the proposition that individuals pass acquired traits on to their offspring. Second, learners answered questions designed to lead them through the process of reinventing the theory of natural selection. Lastly, students participated in the Stebbins and Allen simulation. Geraedts and Boersma reported that this strategy was effective for the majority of students, with 72 percent of students, aged 15 to 16, developing the intended neo-Darwinian theory or a Darwinian theory of natural selection (Geraedts & Boersma, 2006).

Although the original design and subsequent variations of the Stebbins and Allen simulation did not expressly mention the constructivist approach, many of the elements fall in line with the model:

1. *Dissatisfaction*: By using non-living objects (paper chips) as simulated organisms, this simulation directly challenges many of the popular misconceptions listed previously in this proposal. For example: Paper chips are unable to want/need/plan. Paper chips are unable to change colors within a generation. Paper Chips are not able to directly use any of their features. The selection process alone never creates new or additional colors. Asking a series of introspective questions about these concepts can lead students to become dissatisfied with their current conceptions.
2. *Intelligible*: The process used in this simulation is simplified to the point of being easy for students to comprehend.
3. *Plausible*: Seeing the results of the simulation first-hand provides students with a strong sense of plausibility that the model can bring about the results they just observed.
4. *Fruitful*: An accurate theory of evolution explains how the paper chips in the simulation show a change in population.

Using Technology to Improve the Simulation's Effectiveness

Despite being a popular simulation, one notable aspect of the Stebbins and Allen (1975) simulation is that it can be difficult, time consuming, and even frustrating to perform. Carrying out the entirety of the simulation requires multiple tasks, such as the creation, organization, and distribution of paper chips and fabric habitats, as well as the counting, sorting, and redistribution of the next generation of chips. Many of these tasks are not germane to the core concept of natural selection. Instead, they take up attention and cognitive load that could otherwise be used by students to better understand and transfer the material.

Likewise, there is a propensity for error that comes from chips being lost, overlooked, miscounted, or miscalculated. Attention must be directed to preventing, accounting for, or ignoring these errors. This, too, results in extraneous cognitive load to the learner and may detract from learning.

For this study, a tablet-based, digital version of this simulation was developed to alleviate these difficulties. In this tablet-based version, virtual representations of the paper chips and multicolored environment are displayed on the screen of a mobile device. Students use the touch screen interface to perform the "capture" actions of the simulation, and the simulation itself tracks, counts, and displays the results of each generation.

The digital version of the simulation also alleviates many of the possible errors that may occur in the physical version of the simulation. Digital chips cannot be blown by the wind, do not accidentally stick to students' arms and elbows, and cannot be overlooked or lost during the counting and distribution phases. Likewise, students do not need to focus their attention on drawing accurate graphs of the correct size and proportion; the software handles this for them.

This digital version of the simulation is also helpful for teachers who hope to use the simulation in a classroom environment. Computer-generated paper chips do not need to be punched, sorted, or distributed by students or an instructor. The colors of the digital chips are drawn directly from the environment image’s color data, relieving instructors of the challenge of finding or creating chips that match the habitat.

With the popularity of the use of AR in education growing rapidly, an AR version of the simulation was also developed. This digital simulation runs on an Android tablet and uses a “Window-on-the-World” style of AR, using the rear-facing, pass-through camera to provide the illusion of students being able to "look through" the device and into an augmented version of reality. Like the tablet-based digital version, students use the touch screen interface to perform the "capture" actions of the simulation, and the simulation itself can track, count, graph and display the results of each generation of actions. A large, table-sized (26 inch by 20 inch), color photograph of flowers acts as both the habitat for the purposes of the simulation and as a visual marker for the planar surface detection of the AR system.

This AR version of the simulation was hypothesized to provide a number of advantages over a non-AR version. Like the original, paper-chip simulation, the environment is visible in the real world, without the use of a mobile device; users can move to different sides of the environment image and look at the image from different angles both in and out of the technology, allowing participants to make a spatial connection between the physical space in the real world and the digital space in which the prey entities exist. Likewise, participants in an AR version of the simulation can move the mobile device around the environment, looking at the prey entities from different points of view. This movement may likewise provide a spatial connection, and may also provide a sensory-motor connection, both of which may allow users to draw mental connection between the simulation and their real-world experiences. Testing the hypothesis that these advantages provided by an AR version of the simulation are more effective in alleviating the complex misconceptions surrounding natural selection was the foundation for the first question in this study:

> **Research Question 1:** Does an AR version of a popular simulation alter the number of misconceptions expressed by students of natural selection, relative to a non-AR version of the simulation?

Symbolic Versus Representational Visuals in AR

In November of 2017, Google released Google Poly, a creative-commons licensed library where AR developers can browse and download free 3D objects and scenes for use in their projects. By making a wide variety of virtual objects available for developers to freely use, Google intends these art assets to become a uniform standard for building for Augmented Reality and Virtual Reality projects (Zvinakis, 2017).

This library contains numerous 3D models that could be used to make this AR simulation more representative of a real-world, concrete example. For example, the Google Poly library contains multiple versions of a ladybug beetle that could be implemented to transform the simulation from a purely symbolic version using colored circles to represent prey to a more concrete representation of these insects and thus make the simulation more representational of natural selection in the real world.

This concept of using virtual objects that are representative of specific, representational, real-world objects is in alignment with other popular representations of educational AR, such as Microsoft's HoloLens (https://www.microsoft.com/enus/hololens) and Magic Leap (http://www.businessinsider.com/incredible-augmented-reality-headset-2016-3), which focus on the overlay of high-definition, often realistic, virtual objects over the real world. Part of the allure of AR technologies is the ability to display virtual versions of real objects that would otherwise be difficult to present in a classroom environment because of their size, cost, rarity, or imperceptibility.

However, the literature has shown that the use of representational examples, as opposed to generic, symbolic instantiations, can have a negative effect on learning and transfer. Studies have shown that the use of representational examples, especially in the areas of math and science, can detract from the learning experience (Fyfe, McNeil & Goldstone, 2014; Kaminski, Sloutski & Heckler, 2008). Likewise, the coherence principle of multimedia development, which has been supported by numerous studies, states that learners learn better when only coherent information be presented, and recommends that additional information, including details that are interesting but unnecessary, referred to as "seductive details," should be excluded from instructional design (Mayer, 2005). These unnecessary details have been shown to be detrimental to learning by priming an inappropriate base of prior knowledge for the user. This inappropriate prior knowledge may be erroneous information or an inappropriate schema for organizing the material. They can also confuse learners as to what a particular lesson is about (Harp & Mayer, 1998). This is especially relevant to the AR simulation, as the problem of inappropriate prior knowledge is precisely what the AR simulation is intended to resolve.

However, further studies have shown that the optimal approach to seductive details is not as simple as always avoiding them. Research has found that seductive details are not problematic under all conditions. If learners have invested enough cognitive resources to have the resources free to integrate the seductive details, these details are not detrimental. For this reason, some modalities can compensate for the detrimental effect of seductive details (Park, 2015).

In order to test if AR simulations show this compensatory effect, or if there is a difference between best practices recommended by the scientific literature on multimedia design and the recommendations and marketing of commercial entities such as Microsoft, Magic Leap, and Google, a second version of the AR simulation was developed. In this second version, representational objects — ladybug beetles from the Google Poly library — take the place of the symbolic paper-chips used in the simulation.

The literature shows that in the case of traditional desktop computer simulations, the addition of icons that change a simulation from a purely symbolic simulation to a representative one, such as adding animated flames to represent a heat source or using the image of a pressure gauge to display pressure, does not result in a significant increase in the comprehension of the material or the transfer of the material to new situations (Plass et al., 2009). It was hypothesized that this study would confirm these results could be applied to AR simulations. For this reason, this study was limited to only comparing differences between the two AR-based conditions. It was anticipated that AR would be found to be one of the modalities capable of compensating for the detrimental effects of seductive details, and thus it was hypothesized that there would be no significant difference in the change in the total number of misconceptions expressed between the two AR-based conditions.

The way that entities are visualized in the simulation may play a more direct role in at least one of the misconceptions chosen for this study. The misconception referred to as intentionality, that evolution occurs as the result of human-like intention or desire expressed by the entities involved, may be directly affected by how the entities are visualized in the simulation. A review of the research shows that the concept of humans inferring intentionality in inanimate objects is well studied (Scholl & Tremoulet, 2000). Stimuli as simple as a few, small-moving 2-D geometric shapes displayed on a computer screen can be interpreted to have properties of causality, animacy, and intentionality. This phenomenon of inferring intention emerges early in life and is automatic.

A study by Tang, Biocca, and Lim (2004) found that ratings of naturalness for AR were not significantly higher than ratings of naturalness in VR. Even though the AR environment is literally the real-world environment, participants in their AR-based condition reported the environment as less real by than participants in a VR-based condition. Tang et al. theorized that the juxtaposition of computer-generated graphics over the real-world results in participants finding the environment less natural and less believable than a purely virtual environment, such as VR or other non-AR simulation. They theorized that the contrast between the obviously computer-generated images and the real-word environment made all elements feel less natural and less believable.

Other studies have found that believability, such as the perception of AR or symbolic entities as less real, has a direct, neurological connection to the misconceptions studied in this experiment, especially the misconception of intentionality. Mar et al. (2007) found that parts of the brain identified as being active during observations of features that cue intentionality were more active while participants were viewing realistic depictions of social interactions than while participants were viewing cartoony depictions of the same agents and events. That is, when participants viewed realistic images of social interactions, parts of the brain used to determine whether the observed agent is acting intentionally were activated. This is true even when participants were not instructed to look for intentionality in the depictions.

Research Question 2: Does the type of representations—from representational objects (ladybugs from the Google Poly library) to symbolic objects (colored circles) —used in an AR simulation differentially impact the number of misconceptions expressed by students of natural selection?

The ARCS Model of Motivation

The role of motivation as a key element of student learning is an important concept in the study of why some students succeed in an educational context, while others may not be as successful (Pintrich, 2003). The ARCS model of Motivational Design is a commonly used model for designing motivational instructions in multiple educational environments and modalities. The ARCS model gets its name from the four constructs on which it focuses: attention, relevance, confidence, and satisfaction. According to Keller (2010, pp 44-46), these constructs are necessary for learners to be motivated to learn and continue to learn after a learning activity has started.

- *Attention*: Learner’s curiosity and interest must be stimulated and sustained.

- *Relevance*: Learners must feel a connection to the learning material. Learners must believe that the instruction is related to important personal goals or motives.
- *Confidence*: Learners must feel that they are capable of learning from the instructional material. Barriers to confidence not only include fears that the learner may lack the skills to understand or retain the information, but also includes the learner incorrectly believing that they already know the material and thus would not learn anything new from the instructional material.
- *Satisfaction*: To continue to be motivated to learn, learners must have feelings of satisfaction with the process or results.

The four factors of the ARCS model of motivation have been studied in a wide variety of educational contexts. Although the majority of these studies focused on the affective responses of participants, many studies related to student achievement, learning gains and retention were performed. The results of these studies were inconsistent, while some showed no differences, others showed a significant increase in achievement, learning gains, and retention when the ARCS model was applied (Li & Keller 2018).

To measure perceived motivations based on the ARCS model, Keller (2010) developed a 36-item self-reported survey, the Instructional Materials Motivations Survey (IMMS), which measures people's scores on the attention, relevance, confidence, and satisfaction constructs, as well as providing a cumulative overall motivation score. The IMMS was then tested for validity and reliability. To test for validity, participants in a control group were presented with a lesson prepared according to standard principles of instructional design. Participants in the experimental group were presented with a lesson that was enhanced with the intent of increasing participant perception of attention, relevance, confidence, and satisfaction. IMMS scores for the experimental group were significantly higher than scores for the control group. In a separate study, the survey was administered to students in two classes, and the estimated reliability was found to be satisfactory.

The Reduced Instructional Materials Motivation Survey (RIMMS) is a 12-item version of the IMMS questionnaire. The RIMMS has been validated by the literature and has been found to accurately measure the four constructs of the ARCS model (Loorbach, Peters, Karreman & Steehouder, 2015). The RIMMS also provides additional benefits of having no reverse-coded items, and the shorter length reduces biases caused by fatigue or boredom.

All three of the treatments were designed to be motivating and engaging, using bright colors and modern technology. All three conditions were presented to the participants under the same premise of being a fun, engaging educational simulation. Therefore, it was hypothesized that motivation would not differ significantly between conditions.

Time on Task as a Measure of Educational Quality

One notable aspect of the Stebbins and Allen (1975) simulation is that, although all students complete the same steps, the time it takes for an individual student to complete the task can vary. Some students may be faster at finding their simulated prey than others. Some may take time to stop and think about the material as they complete the task. The literature has shown that increased time on task is correlated with increased learning, specifically in cases of computer-based training where time on task is allowed to vary between participants (Brown, 2001). By viewing time on task as an important user choice, it has been theorized that many participants who finish quickly may skip elements that are critical to learning. Some Serious Games researchers have even gone so far as to state that time on task, driven by motivation, is the most influential factor in student achievement, and one of the primary motivations for including games and digital simulations in education (Annetta, Minogue, Holmes, & Cheng, 2009). For this reason, time on task was compared between the contrasts. Although the simulation did not force participants to complete the tasks at a predetermined rate, it was hypothesized that time on task would not vary between conditions, confirming that the results of this study are not simply due to differences in time on task.

Chapter 2

Method

Participants

The participants were 189 undergraduate college students attending Arizona State University, a research-intensive university in Tempe, Arizona with a population of about 100,000 students. A short demographic survey was issued to track age, self-reported gender, and degree of study. Participants ranged in age from 18 to 48, with a mean age of 20 years old. Of the 189 participants, 94 (49.74%) were male, 95 (50.26%) were female. Measures were also included for previous exposure to AR technology, use of tablets/smartphones and other similar devices, and the participants' self-reported level of familiarity with the topic of natural selection.

The majority of participants reported being familiar or very familiar with computer-based education (129, 68.25%), mobile devices (186, 98.41%), and natural selection (157, 83.07%). However, a majority of participants reported being "not familiar" or "neither familiar or not familiar" with Augmented Reality (153, 80.95%). The majority of participants also reported having a "very positive" or "positive" attitude towards Biology (118, 62.43%), Natural Selection (114, 60.32%), and Mobile Devices (175, 92.59%). Participants also tended to self-report their individual skill level with mobile devices at a level of Excellent or Good (184, 97.35%). However, the majority of participants rated their education level in Biology as Fair (107, 56.61%) or Good (59, 31.22%), with only 8 (4.23%) self-reporting their education level as Excellent, and only 15 (7.94%) reporting their education level as Poor.

Participants were invited to participate in this study through multiple methods, with the majority of participants recruited through the SONA web-based participant pool system in the education department at Arizona State University. Announcements were also made to various classes. Participants were from a variety of majors and received course credit in exchange for their participation. Participants were allowed to choose from available time slots based on researcher availability. Available timeslots ranged from 10:30 am to 6:30 pm and included both weekends and weekdays.

All participants were over the age of 18 years of age, and no identifying information was attached to the collected data. Therefore, an Institutional Review Board exemption status was granted for this study. No participants withdrew from the study and all submitted surveys and assessments were submitted as complete and included in the data analysis. One extreme outlier was found in the measurement of time on task and was excluded only from the time on task analysis.

Design

This study's design consisted of a between-subjects design with a single independent variable consisting of three levels. Specifically, participants were randomly assigned in equal numbers to one of three conditions:

1. A Non-AR Representational Simulation condition which used the tablet-based adaptation of the Stebbins and Allen simulation, with virtual "prey" represented by 3D models of ladybug beetles from the Google Poly Library;
2. A Representational AR condition which used a "window-on-the-world" Augmented Reality system, with the same 3D models as the first condition; and

3. A Symbolic AR condition which also used the "window-on-the-world" AR technology. However, the virtual "prey" to be collected was represented by colored circles.

Learning Context

Environment. The study took place in a small workspace-style environment. Up to three participants were able to take part in the study at a given time. Participants were seated at tables located along the outside of the room, facing the walls, and were unable to see the screens of the other participants. Each table had a large (26 inch by 20 inch) printed copy of a photograph of flowers attached to the table.

Hardware / Mobile Device. The software was run on a Lenovo TAB4 8 Android Tablet with a Qualcomm Snapdragon 425 CPU, and an 8.0-inch 1280x800 high-resolution display. The tablet OS was Android 7.11 and the rear-facing camera used in the AR simulation provided images at a resolution that matched the resolution of the display (1280x800) at approximately 60 frames per second.

Software. The digital simulation software was developed by the researchers to meet the requirements of the 1975 Stebbins simulation description. The software was built in the Unity game engine (www.unity3d.com) version 2017.3, and used the Vuforia API for AR that is packaged with that version of the engine.

Treatment Conditions

All Conditions. In all treatment groups, the simulation closely follows the methodology presented in the Stebbins and Allen (1975) simulation. As with the original simulation, one hundred "prey" entities were placed on a simulated environment. Although these prey entities were placed using the pseudorandom number generator in the Unity engine, the generator was seeded with the same value for all participants, ensuring that the entities were placed in the same relative virtual locations for all participants. All conditions used the same virtual environment, a full-color photograph of flowers. This habitat image was predominately purple in color, with some green and yellow areas. In all conditions, the prey entities were recolored to create a variety of different colored prey, some of which used exact purple, green, and yellow colors pulled from the habitat image. In all conditions, participants tapped on prey to "capture" them, and the time elapsed and number of prey entities captured were shown on the screen in the same way. All conditions used the same priming questions and methods for displaying generational data and changes.

Figure 1. Habitat image, all conditions.

Non-AR (Representational) Condition. In the tablet-based, Non-AR treatment, the virtual habitat used in the simulation was entirely virtual. That is, although a printed version of the habitat was attached to the participant's study area, objects from the real-world were not registered or displayed by the software and did not play a direct role in the simulation. The prey entities used in this condition were representational 3D models from the Google Poly library. The beetles' elytra (shells) were recolored in the same way that paper chips or color circles would have been presented.

Figure 2. Tablet-based Non-AR condition (closeup)

Figure 3. Tablet-based Non-AR condition (in action)

Symbolic AR Condition. In this treatment group, a window-on-the-world AR environment was used to display the virtual data onto the real world. The rear-facing camera of the tablet acted as a pass-through camera to display a live image of the real-world behind the device. On top of this camera image, the virtual symbols that represented the prey entities were displayed as colored circles (discs). These circles used exactly the same colors as the other conditions, and were placed in the same virtual locations, relative to the habitat image, as in the other conditions.

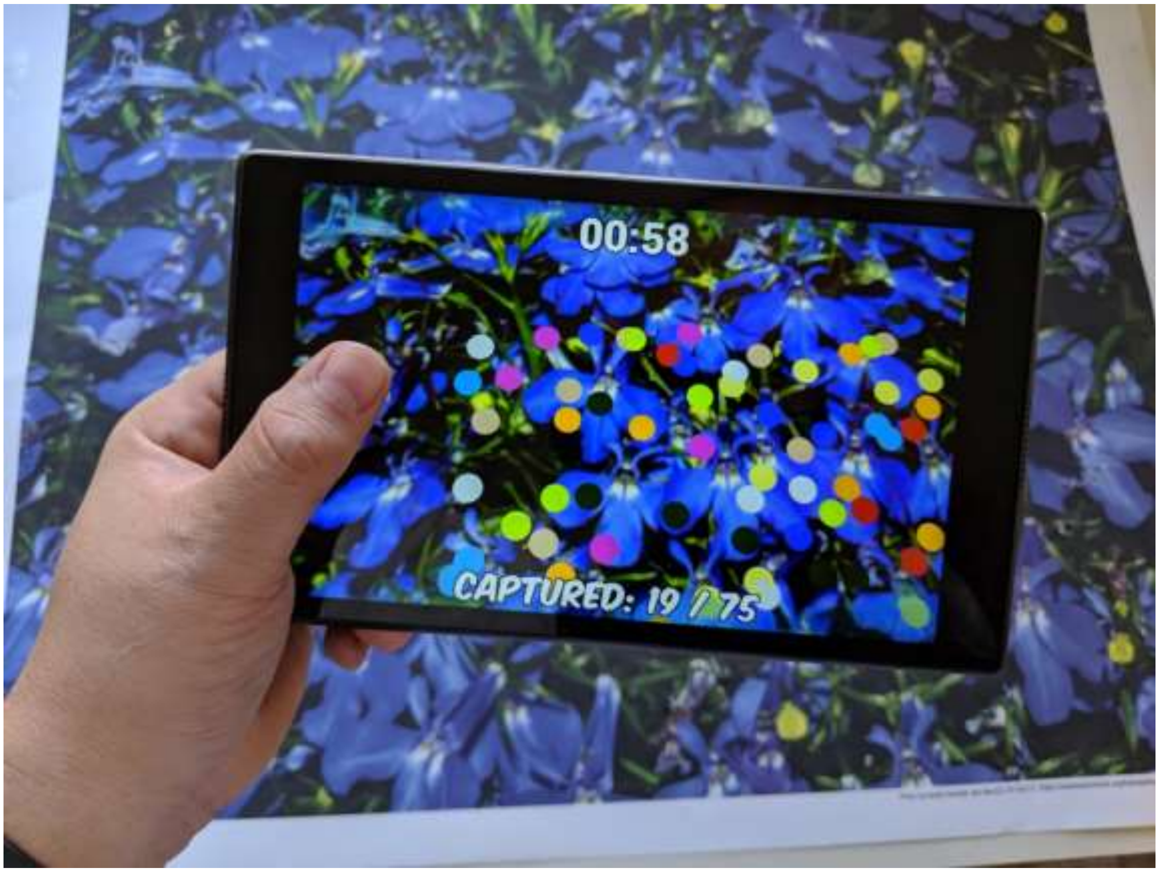

Figure 4. Symbolic AR condition (closeup)

Figure 5. Symbolic AR condition (in action)

Representational AR Condition. The simulation used by this treatment group also used a window-on-the-world style AR environment. However, this simulation again used the recolored ladybug beetle models from the Google Poly library. Apart from this cosmetic change, the two AR conditions are otherwise identical.

Figure 6. Representational AR condition (closeup)

Figure 7. Representational AR condition (in action)

Instruments

Two instruments were administered before the treatment: a background questionnaire (see Appendix A) and pretest (see Appendix C); and two instruments were administered after the treatment: a posttest (see Appendix C) and user experience questionnaire (see Appendix B). These instruments were administered as printed, pen-and-paper assessments. Participants completed the surveys, including the pretest and posttest, in the same environment as they received the treatment. Each of the instruments are described in more detail below.

Background Questionnaire. Participants completed a written questionnaire which included the following elements:

- *Demographics*: gender, age, ethnicity, major, and number of completed college semesters.

- *Familiarity and Attitude Towards Topic*: self-rating of participants' self-reported skill level in Biology (excellent, good, fair, poor), as well as their attitude towards the topic (very positive, positive, neutral, negative, very negative.), self-reported familiarity with natural selection (very familiar, familiar, neither familiar or not familiar, not familiar), and their attitude towards natural selection (very positive, positive, neutral, negative, very negative.)
- *Familiarity and Attitude with Mobile Devices and AR*: self-rating of participants' self-reported experience with computer-based education, mixed and augmented reality, and mobile devices (very familiar, familiar, neither familiar or not familiar, not familiar, no experience), as well as their attitude towards mobile devices (very positive, positive, neutral, negative, very negative), and their self-reported ability to use mobile devices (excellent, good, fair, poor.)

Participant data was only identifiable by a unique participant id number.

Pretest and Posttest. After completion of the presurvey, but prior to treatment, an assessment was administered to participants. This 10-question pretest began with three short-answer questions to prime participants to think deeply about the topic of natural selection. In the remaining seven questions, participants were presented with descriptions of biological phenomena and asked to choose one or more statements about that phenomenon that are true in terms of how evolutionary biologists use and understand the theory of evolution today. Each question was accompanied by five possible statements. Participants were instructed to choose all statements they believe to be correct. These five statements were chosen to express each of the four misconceptions analyzed in this study and the fifth statement was a correct interpretation of evolutionary theory. The order in which the statements (answers) appeared was randomized when the test was developed but was the same for each participant.

After the completion of the treatment, the same instrument was administered as a posttest to measure any change in the number of misconceptions expressed by the participants. The only difference between the pretest and posttest is the addition of an 11th question in the posttest. This final question asked students to think back on the simulation, describe any patterns they may have seen, and provide a possible theory as to why that pattern may have occurred.

Scoring of the Pretest and Posttest. Participant answers on both the pretest and posttest were scored on the number of misconceptions expressed on the seven multiple answer questions. The three priming questions and the additional posttest question were not scored. Each of the four misconceptions were scored, with the total number of times that misconception was scored across the entire test recorded, with a lowest possible score being zero if the misconception was not expressed in any of the answer, and a seven if the misconception was expressed in all seven questions. The total number of times the correct answer was selected was also recorded, but an inverted scoring method was used. The lowest possible score for the correct answer being zero if the correct statement was selected in all seven questions. The highest possible score being seven if the correct statement was not selected in all seven questions.

The total number of misconceptions for the entire assessment was also recorded. For each individual question, a five was the highest possible total, given for an answer that selected all four of the misconceptions identified for this study, but did not select the correct interpretation. This allowed a single assessment a maximum cumulative total of

35, five points each for the seven multiple-choice questions. The entire pretest and posttest instruments are available as Appendix C.

Post-Activity Survey. After the treatment and posttest, a written questionnaire was administered as a measure of participant motivation. This survey was designed based on the RIMMS survey developed by Loorbach et al. (2017). However, key terms related to the topic and methodology were changed. For example, where the original RIMMS asked participants to rate the trueness of the statement "The content and style of writing in these user instructions convey the impression that being able to work with the telephone is worth it", this statement was adapted to work in the context of the simulation as "The content and style in this lesson convey the impression that being able to use the information presented is worth it." Like the other measures, this questionnaire was a pen and paper assessment.

Scoring of the Post-Activity Survey. The RIMMS questionnaire contains twelve Likert scale style statements that can be organized into three questions per element of the ARCS model of motivation. For each statement, a score of one to five was recorded representing answers of not true, slightly true, ,moderately true, mostly true, and very true, respectively. The total values for each element of the ARCS model were recorded. Likewise, the total value of all responses was recorded as a measure of total motivation.

Time on Task. The following times were manually recorded by the researcher for each participant. The time the participant began the demographic survey, the time the participant began the pretest, the time the participant began the treatment, the time the participant began the posttest, the time the participant began the motivation survey, and the time the participant submitted the motivation survey were all recorded. There was no down-time between stages, so these times were used to determine the duration and time on task for each stage in the experiment. The difference between the time the participant began the treatment and the time the participant began the posttest was recorded as the treatment time and used in all time on task calculations.

Procedure

Registration and Random Assignment with Equal Numbers. Participants registered for the study through an online registration form. Participants then selected a time from a list of time slots made available based on the researcher's availability. Up to three participants were allowed to register for each time slot. Upon arrival at the study location, participants were allowed to choose from one of four available workstations and were assigned a random condition. Each of the workstations contained a chair, a table, and a large (26 inch by 20 inch), printed photograph of flowers that would act as an AR marker and environment for the AR-based conditions. The informed consent form was present at each workstation, along with a pen for completing the written assessments.

To ensure group sizes remained balanced, the first participant was assigned a random condition out of the three conditions. The second participant was assigned a random condition from the remaining conditions. The third participant was then assigned the final, remaining condition. This process was repeated, with the fourth participant being randomly assigned to one of the three conditions, and so on, until all participants had completed the study. Although over 200 participants registered through the online system, only 189 participants participated in the study, which allowed for an equal number of participants, 63, in each condition. Table 1 shows sample size by condition.

Table 1

Sample Size by Condition

	Non-AR	AR-Symbolic	AR-Representational
Sample Size	63	63	63

Informed Consent and Presurvey. After being seated at their self-selected workstation, participants read and signed an informed consent form, agreeing to participate in the study. As the participants completed the consent form, the researcher prepared the tablet device for the condition assigned to the participant. Upon submitting the consent form to the researcher, participants were given a written pretreatment demographic survey and presented with the tablet device. The software on the tablet contained a screen instructing the participant to complete the presurvey before advancing to the next step. The researcher informed the participants that they would be using the tablet later in the study, but for the moment it was being used to track their status in the study process. The software required participants to wait at least 30 seconds before a next button in the bottom-right corner of the screen would become active and allow participants to advance from the survey to the pretest. However, all participants took longer than 30 seconds to complete the presurvey and no participants had to wait for the button to become active.

Pretest. After completing the presurvey, participants handed the survey to the researcher and in return were presented with the written pretest. The pretest, shown in Appendix C, consisted of 10 questions. The first three questions were short-answer questions designed to ensure participants were thinking deeply about natural selection and evolution before answering the measured questions. The remaining seven questions were multiple choice questions which presented participants with an example of a natural phenomenon and asked them to indicate all of the statements likely to be true in terms of how evolutionary biologists use and understand the theory of evolution today. Like the presurvey, participants were unable to advance the tablet software from the pretest screen to the treatment until after at least 30 seconds had passed. No participants completed the pretest in less than 30 seconds; therefore, no participants had to wait for the next button to become active.

Treatment - Tutorial. After completing the pretest, participants were able to advance the software to the simulation activity for their condition. A short tutorial on how to complete the activity was available on the tablet for all conditions. First, participants were welcomed to the Natural Selection Simulation and instructed in how to use the next button to advance between sections.

Next, participants were informed that they were to act as predators capturing and eating prey, and to do so, they were to tap on the prey they wish to capture. They were shown an example of an animated hand tapping a prey object. The appearance of the prey objects, in this and all other screens, was dependent on the participant’s assigned condition. Non-AR and Representational-AR conditions were presented with 3D models of ladybug beetles, while participants in the Symbolic-AR condition were presented with a colored circle. All prey creatures in the tutorial section were red in color.

In the next step of the tutorial, participants were instructed to tap on sample prey creatures that were displayed against a dark blue background. Participants were unable to advance until they “captured” all three prey.

After capturing prey on the solid background, participants were shown a new environment and instructed to capture three additional prey in the new environment. In the non-AR condition, this environment was an on-screen image of the flower photograph present at the workstations. In the AR-based conditions, this new environment was an image from the rear-facing camera. Participants were instructed to "Take a moment to look at the flower environment on [their] device. If [they] can't see any prey, try looking at a different part of the flower environment." Participants could not advance to the next section until they captured the three prey in the new environment. After capturing all three prey, the next button became active and participants could advance to a screen instructing them to hunt swiftly and capture 75 prey in the simulation as quickly as possible.

Treatment – Simulation Activity. Before each capture session of the simulation, a 3-second countdown timer instructed participants to "Get Ready." After the timer, they were presented with the simulation. One hundred prey entities were placed in the virtual environment. As with the tutorial, the environment in the non-AR condition was a single image of the flower photograph. In the AR-conditions, the environment background was rendered from the rear-facing camera of the device. The appearance of the prey entities was also dependent on condition. In the Symbolic-AR condition, the prey entities were displayed as solid-colored circles. In the Non-AR and Representational-AR conditions, the prey entities were represented as 3D models of ladybug beetles from the Google Poly library.

A timer at the top of the screen tracked how long the participants were in the current simulation mode and was designed to create a sense of urgency. At the bottom of the screen, the number of prey entities captured, out of the goal of 75 captured, was presented to the participant. Participants would tap on prey creatures to capture them. Each capture was accompanied with a visual particle effect and a crunch sound effect. A delay timer required participants to wait at least 0.5 seconds between taps, preventing participants from simply mashing their hand or other body-part on the screen to capture multiple prey at once.

After 75 prey were captured, the simulation automatically advanced to a congratulatory screen that required participants to click a next button to advance. The participants were then presented with a screen that systematically displayed all of the prey entities that were not captured by the participant, sorted by color. These 25 uneaten prey entities were organized in four rows. After these prey entities were displayed to the screen, the next button became active.

After pressing next, participants were informed that "Offspring are created from each survivor," and the software systematically added three new, same-colored entities directly below each uneaten prey entity. The next button then became active again, and after the next button was pressed, participants were informed that "the prey returns to the environment" as the prey entities were removed one-at-a-time from the screen. After all prey objects were removed, the next button became active again.

Pressing the next button presented the participants with a "Think About It Screen" that asked questions designed to lead participants to think about the desired learning material. Participants were required to remain on this screen for 30 seconds before the next button became available. Pressing the next button then returned the software to the countdown and the capture session of the next generation of the simulation began.

Participants repeated these simulation steps through three generations of prey entities. Each generation was accompanied by a different set of “Think About It” text prompts. These prompts are detailed in Table 2.

Table 2.

“Think About It” Text

Generation	Text
1	A “gene” is a unit of heredity that transfers a trait from one generation to the next. What are the genes and traits of the prey you are capturing?
2	Did any prey change colors during the game? Did any individual prey ever change its genes?
3	Do the prey objects have any needs or desires? Do they make decisions?

After completing three generations, participants were then presented with a summary screen and asked if they identify a pattern and why do they think this happened. The summary screen displayed all 100 prey entities from each generation. This allowed participants to see how the color of prey entities changed across generations. Participants were required to wait 30 seconds on this screen before the next button became active and they could advance to the next screen.

Figure 8. Final summary screen

Posttest. After completing the simulation, participants were instructed to complete the posttest. The first 10 questions of the posttest were identical to the pretest. Question 11 of the posttest, however, asked participants if they saw a pattern and to describe why they thought that pattern occurred.

Post-Treatment Survey (RIMMS). After completing the posttest, participants were presented with a written survey based closely on the RIMMS survey created by Loorbach, Peters, Karreman & Steehouder (2015). Only words and phrases that directly related to the methodology and topic at hand were changed, such as changing the phrase “on the pages” to “in the simulation” and “work well with the telephone” being generalized to “work with the material presented”. This survey is included as Appendix B. After completing the Post-treatment survey, participants were released.

Chapter 3

Results

Introduction

Two research questions were analyzed using a priori orthogonal contrasts in this study. To answer these questions, several variables were studied for each of the conditions. Table 3 lists each of the questions and details the analytical approach used for each question.

Table 3

Data Sources and Analyses

Research Question	Data Sources	Analysis
1. Does an AR version of a popular simulation alter the number of misconceptions expressed by students of natural selection, relative to non-AR version of the simulation? a. Does an AR version of a popular simulation alter the motivation of users, relative to a non-AR version of the simulation? b. Does an AR version of a popular simulation alter the time spent interacting with the simulation, relative to a non-AR version of the simulation?	Pretest Posttest RIMMS survey Time on task in the simulation.	t-tests with planned orthogonal contrasts
2. Does the type of visualization—from representational (ladybugs from the Google Poly library) to symbolic (colored circles—used in an AR simulation differentially impact the number of misconceptions expressed by students of natural selection?	Pretest Posttest RIMMS survey Time on task in the simulation.	t-tests with planned orthogonal contrasts

These planned research questions were analyzed with a pair of orthogonal contrasts, which were designed specifically for testing the nonredundant and independent a priori hypotheses embodied by these research questions. One contrast compared the mean number of misconceptions held by students who participate in the tablet-based, non-AR simulation to the mean number of misconceptions held by students who participate in either of the two AR simulations (symbolic and representational). The second contrast was used to compare the mean number of misconceptions held by students who participate in the representational version of the AR simulation from the mean number of misconceptions held by those who participate in the symbolic version of the AR simulation. Since these planned comparisons meet the requirements for orthogonality, t-tests were used to analyze each one with an α of 0.05. Table 4 below captures the null hypothesis for two contrasts corresponding to the research questions and provides the contrast weights used in the analysis.

Table 4.

Contrasts

Contrast / Treatment	**Representational Non-AR Simulation (μ_1)**	**Representational AR Simulation (μ_2)**	**Symbolic AR Simulation (μ_3)**
Research Question 1: H_0: $\mu_1 = \frac{\mu_2+\mu_3}{2}$	1	-½	-½
Research Question 2: H_0: $\mu_2 = \mu_3$	0	1	-1

Pretest and Posttest Scores

The pretest scores for each misconception, the correct statement, and the overall number of misconceptions expressed on both the pretest and the posttest were recorded. In the case of the four misconceptions, the score represents the number of times the misconception-based answer was selected. In the case of the correct answer, the scoring was inverted, with the number representing the number of times the correct answer was not selected. The total misconceptions score represents the total of all the misconceptions as well as the inverted score for the correct answer. The pretest scores are presented in table 5 and the posttest scores are presented in table 6.

Table 5.

Mean Scores and Standard Deviations on Pretest by Misconception

Misconception	Condition	N	Mean	SD
	Non-AR	63	4.43	1.78
Teleological	AR Symbolic	63	4.57	1.76
	AR Representational	63	4.87	1.45
	Non-AR	63	1.95	1.21
Soft Inheritance	AR Symbolic	63	2.32	1.64
	AR Representational	63	2.17	1.55
	Non-AR	63	2.68	1.71
Intentionality	AR Symbolic	63	2.73	1.64
	AR Representational	63	2.75	1.37
	Non-AR	63	2.44	1.39
Event-Driven	AR Symbolic	63	2.25	1.49
	AR Representational	63	2.56	1.56
	Non-AR	63	3.02	1.73
Correct Answer	AR Symbolic	63	2.51	1.81
	AR Representational	63	2.68	1.86
	Non-AR	63	14.52	4.42
Total Misconceptions	AR Symbolic	63	14.38	5.22
	AR-Representational	63	15.03	4.86

Table 6.

Mean Scores and Standard Deviations on Posttest by Misconception

Misconception	Condition	N	Mean	SD
	Non-AR	63	4.62	2.08
Teleological	AR Symbolic	63	4.43	2.13
	AR Representational	63	4.56	2.36
	Non-AR	63	2.68	1.63
Soft Inheritance	AR Symbolic	63	2.49	1.97
	AR Representational	63	2.68	1.83
	Non-AR	63	3.63	1.99
Intentionality	AR Symbolic	63	2.84	2.26
	AR Representational	63	2.81	1.82
	Non-AR	63	2.98	1.71
Event-Driven	AR Symbolic	63	2.62	1.92
	AR Representational	63	2.75	1.75
	Non-AR	63	2.62	1.77
Correct Answer	AR Symbolic	63	2.24	1.72
	AR Representational	63	2.60	1.98
	Non-AR	63	16.54	6.13
Total Misconceptions	AR Symbolic	63	14.62	6.71
	AR-Representational	63	15.40	6.45

Change in Total Number of Misconceptions Expressed

The number of misconceptions expressed in the pre and post assessments were measured and the change in the number misconceptions expressed (posttest count minus pretest count) was calculated. The mean and standard deviation for each condition is presented in Table 7. Positive numbers show an increase in the number of misconceptions expressed.

Table 7.

Mean Scores and Standard Deviation of Change in Number of Misconceptions Expressed

Condition	N	Mean (Gain)	SD (Gain)
Non-AR	63	2.02	4.45
AR-Symbolic	63	0.24	3.68
AR-Representational	63	0.37	3.80

A t-test using the contrasts from Table 4 was used to evaluate the effects of the use of AR on the change in number of misconceptions expressed by participants. A significant difference was discovered in the change of the number of misconceptions expressed by participants in the tablet-based, non-AR condition compared to the AR-based Symbolic and AR-based Representational conditions; $t(186) = 2.785$, $p = 0.006$. Cohen's d was used as an effect size index, where 0.20, 0.50, and 0.80 represent small, medium, and large effect sizes, respectively. A small-to-medium effect size of $d = 0.408$ was found.

A t-test using the contrasts from Table 4 was used to evaluate the effects of the type of entities used in the simulation, symbolic compared to representational, on the change in number of misconceptions expressed by participants. No significant difference was discovered in the change of the number of misconceptions expressed by participants in the AR-based Symbolic condition compared to the AR-based Representational condition; $t(186) = -0.179$, $p = 0.858$. A trivial effect size of $d = 0.026$ was found.

Change in Number of Expressions by Misconception

The change in the total number of times each misconception was expressed (posttest minus pretest) was calculated. The means and standard deviations for each misconception are displayed in Table 8 below.

Table 8.

Mean Scores and Standard Deviations in Change in Number of Expressions by Misconception

Misconception	Condition	N	Mean	SD
Teleological	Non-AR	63	0.19	1.60
	AR Symbolic	63	0.14	1.80
	AR Representational	63	-0.32	1.73
Soft Inheritance	Non-AR	63	0.73	1.56
	AR Symbolic	63	0.17	1.30
	AR Representational	63	0.51	1.27
Intentionality	Non-AR	63	0.95	1.64
	AR Symbolic	63	0.11	1.60
	AR Representational	63	0.06	1.41
Event-Driven	Non-AR	63	0.54	1.54
	AR Symbolic	63	0.37	1.56
	AR Representational	63	0.19	1.31
Correct Answer	Non-AR	63	-0.40	1.63
	AR Symbolic	63	-0.27	1.17
	AR Representational	63	-0.08	1.46

No significant difference was found in the change in the number of times the teleological misconception was expressed in the tablet-based, non-AR condition compared to the AR-based Symbolic and AR-based Representational conditions; $t(186) = 1.592$, $p = 0.113$. No significant difference was found in the number of times the teleological misconception was expressed in the AR-based Symbolic condition compared to the AR-based Representational condition; $t(186) = 0.572$, $p = 0.568$.

No significant difference was found in the change in the number of times the soft inheritance misconception was expressed in the tablet-based, non-AR condition compared to the AR-based Symbolic and AR-based Representational conditions; $t(186) = 1.824$, $p = 0.070$. No significant difference was found in the number of times the soft inheritance misconception was expressed in the AR-based Symbolic condition compared to the AR-based Representational condition; $t(186) = -1.354$, $p = 0.177$.

No significant difference was found in the change in the number of times the event-driven misconception was expressed in the tablet-based, non-AR condition compared to the AR-based Symbolic and AR-based Representational conditions; $t(186) = 1.152$, $p = 0.251$. No significant difference was found in the number of times the event-driven misconception was expressed in the AR-based Symbolic condition compared to the AR-based Representational condition; $t(186) = 0.665$, $p = 0.507$.

No significant difference was found in change in the number of times the intentionality misconception was expressed in the AR-based Symbolic condition compared to the AR-based Representational condition; $t(186) = 0.172$, $p = 0.864$. However, a significant difference was found in the change in number of times the intentionality misconception was expressed in the tablet-based, non-AR condition compared to the AR-based Symbolic and AR-based Representational conditions; $t(186) = 3.609$, $p<0.001$. A medium effect size of $d=0.529$ was found.

No significant difference was found in the change in the number of times participants in the non-AR condition chose the correctly phrased answer compared to the AR-based Symbolic and AR-based Representational conditions; $t(186) = -1.005$, $p=0.316$. No significant difference was found in the change in the number of times participants in the AR-based Symbolic condition chose the correctly phrased answer compared to the AR-based representational condition; $t(186) = -0.746$, $p=0.457$.

Tests of Aptitude-Treatment Interaction

To test for an aptitude-treatment interaction, participants were grouped into two aptitude categories based on the total number of misconceptions expressed in the pretest. The 21 participants whose score for total misconceptions expressed equaled the median score (15) were removed from the dataset. The 86 participants who scored below the median number of total misconceptions were placed in a high aptitude group. The 82 participants who scored above the median number of total misconceptions were placed in a low aptitude group. The number of participants for each group is presented in Table 9.

Table 9.

Participants per Aptitude Treatment Interaction Group

Condition	High (N)	Low (N)
Non-AR	31	27
AR-Symbolic	27	28
AR-Representational	28	27

A two-way analysis of variance was conducted on influence of condition and aptitude group on the total change in number of misconceptions expressed. No significant interaction effect was found between condition and aptitude group, $F(2,162) = 0.15$, $p = 0.86$. A two-way analysis of variance was conducted on influence of condition and aptitude group on the change in number of times the teleological misconception was expressed. No significant interaction effect was found between condition and aptitude group, $F(2,162) = 0.75$, $p = 0.48$. A two-way analysis of variance was conducted on influence of condition and aptitude group on the change in number of times the soft inheritance misconception was expressed. No significant interaction effect was found between condition and aptitude group, $F(2,162) = 1.73$, $p = 0.18$. A two-way analysis of variance was conducted on influence of condition and aptitude group on the change in number of times the event-driven misconception was expressed. No significant interaction effect was found between condition and aptitude group, $F(2,162) = 2.81$, $p = 0.06$. A two-way analysis of variance was conducted on influence of condition and aptitude group on the change in number of times the intentionality misconception was expressed. No significant interaction effect was found between condition and aptitude group, $F(2,162) = 0.45$, $p = 0.64$. A two-way analysis of variance was conducted on influence of condition and aptitude group on the change in number of times participants chose the correctly worded statement. No significant interaction effect was found between condition and aptitude group, $F(2,162) = 1.18$, $p = 0.31$.

Motivation (ARCS)

The RIMMS survey was used to measure the level of motivation reported by participants in all conditions. Using the RIMMS survey, the participants motivation was measured across the four constructs of the ARCS Model as well as a measure of total motivation, derived from the sum of the four constructs. In the four ARCS constructs, possible values ranged from 3 to 15, with 15 being the highest possible value for the given construct. For total motivation, values ranged from 12 to 60, with 60 being the highest possible value. The mean and standard deviation for each condition is presented in Table 10.

Table 10

RIMMS Motivation Data

Construct	Condition	N	Mean	SD
Attention	Non-AR	63	12.76	2.26
	AR Symbolic	63	12.21	2.37
	AR Representational	63	12.35	2.56
Relevance	Non-AR	63	11.75	2.13
	AR Symbolic	63	12.03	2.50
	AR Representational	63	12.32	2.04
Confidence	Non-AR	63	13.10	2.37
	AR Symbolic	63	13.02	2.09
	AR Representational	63	13.10	2.24
Satisfaction	Non-AR	63	12.57	2.49
	AR Symbolic	63	12.00	2.78
	AR Representational	63	12.63	2.34
Total Motivation	Non-AR	63	49.33	8.52
	AR Symbolic	63	50.32	7.43
	AR Representational	63	49.97	7.67

Using the contrasts from Table 4, t-tests were used to evaluate the effects of AR on the level of motivation reported by participants. Five t-tests were conducted, over for overall motivation and four on the individual elements of the ARCS model. No significant difference was discovered in any of the individual elements of the ARCS model or in the total level of motivation. No significant difference was discovered the expressed levels of attention for participants in the tablet-based, non-AR condition compared to the AR-based Symbolic and AR-based Representational conditions; $t(186) = 1.308$, $p = 0.192$. No significant difference was discovered in expressed levels of relevance for participants in the tablet-based, non-AR condition compared to the AR-based Symbolic and AR-based Representational conditions; $t(186) = -1.245$, $p = 0.215$. No significant difference was discovered in expressed levels of confidence for participants in the tablet-based, non-AR condition compared to the AR-based Symbolic and AR-based Representational conditions; t(186) = 0.389, p = 0.698. No significant difference was discovered in expressed levels of satisfaction for participants in the tablet-based, non-AR condition compared to the AR-based Symbolic and AR-based Representational conditions; t(186) = 0.646, p = 0.519. Lastly, No significant difference was discovered in the cumulative total motivation score for participants in the tablet-based, non-AR condition compared to the AR-based Symbolic and AR-based Representational conditions; t(186) = 0.374, p = 0.709.

A t-test using the contrasts from Table 4 was used to evaluate the effects of the type of entities used in the simulation, symbolic compared to representational, on the level of motivation reported by participants No significant difference was discovered in any of the individual elements of the ARCS model or in the total level of motivation. No significant difference was discovered the expressed levels of attention for participants in AR-based Symbolic condition compared to the AR-based Representational condition; t(186) = -0.334, p = 0.739. No significant difference was discovered in expressed levels of relevance for participants in the AR-based Symbolic condition and AR-based Representational condition; t(186) = -0.719, p = 0.473. No significant difference was discovered in expressed levels of confidence for participants in the AR-based Symbolic condition and the AR-based Representational condition; t(186) = 0.198, p = 0.843. No significant difference was discovered in expressed levels of satisfaction for participants in the AR-based Symbolic condition and AR-based Representational condition; t(186) = -1.399, p = 0.163. Lastly, no significant difference was discovered in the cumulative total motivation score for participants in the AR-based Symbolic condition and AR-based Representational condition; t(186) = -0.718, p = 0.474.

Time on Task

Due to limits on the rate at which participants could capture prey and delays until the next buttons became active, participants in all conditions were equally limited as to the minimum time they were required to interact with the treatment. However, there was no upper limit as to the amount of time participants were allowed to spend with the treatment. The amount of time each participant spent in the treatment was recorded by the researcher manually tracking the time the participant was presented with the treatment condition and the time the participant was presented with the posttest that immediately followed the treatment.

In analyzing the data, a single extreme outlier was discovered. While all other participants completed the treatment between a minimum of 4 minutes 5 seconds and a maximum of 8 minutes 57 seconds (M= 6 minutes 27 seconds, SD= 44 seconds), the outlier spent 20 minutes and 4 seconds in the treatment. For this reason, the outlier data point was removed from the data for the analysis of time on task.

Table 11

Mean Scores and Standard Deviation of Change in Time on Task

Condition	N	Mean	SD
Non-AR	63	6 min 7sec	42 sec
AR-Symbolic	62	6 min 45 sec	41 sec
AR-Representational	63	6 min 30 sec	41 sec

A t-test using the contrasts from Table 4 was used to evaluate the effects of AR on the time participants spent with the treatment. A significant difference was discovered in the time on task for participants in the tablet-based, non-AR condition compared to the

AR-based Symbolic and AR-based Representational conditions; $t(185) = -4.650$, $p < 0.001$. A medium effect size of $d = -0.684$ was found.

A t-test using the contrasts from Table 4 was used to evaluate the effects of the type of entities used in the simulation, symbolic compared to representational, on the time participants spent with the treatment. No significant difference was discovered in the time on task for participants in the AR-based Symbolic condition compared to the AR-based Representational condition; $t(185) = 1.906$, $p = 0.058$. A small effect size of $d = 0.280$ was found.

Chapter 4

Discussion

Findings and General Discussion

One notable element of the results of this study was the direction in which the change in misconceptions occurred. All conditions showed an increase in total misconceptions expressed. However, a significant difference in the increase in total misconceptions expressed was found for users of the tablet-based simulation when compared to users of the AR-based simulations.

When the change in number of expressions for each misconception was analyzed individually, only the misconception of intentionality showed a significant difference in the number of expressions. Like total misconceptions, there was an increase in all conditions. However, participants in the tablet-based simulation showed a larger increase in number of expressions than participants in the AR-based versions. Of the other misconceptions, only the AR versions of the teleological misconception showed any alleviation of misconceptions, however, this change was not significant. There were no significant differences when comparing the use of symbolic virtual entities when compared to representational entities.

Contrary to expectations that the simulation would help to alleviate misconceptions, the treatment resulted in an increase in the total number of misconceptions expressed by participants. Previous studies have found other versions of the simulation to be effective when combined with an external lesson (Geraedts & Boersma, 2006), so it is likely that this implementation of the simulation failed to communicate the elements of the lesson required for the conceptual change to occur in the desired direction. In other words, it is possible that the simulation simply replicated an environment similar to the one in which the participants originally developed their misconceptions. This is consistent with the finding that time on task was significantly greater in the tablet-based condition, the same condition that showed an increase in expressed misconceptions.

Hewson and Hewson (2003) theorized that there are four critical elements required for alleviating misconceptions – Dissatisfaction, Intelligible, Plausible, and Fruitful. These constructivist elements are necessary for conceptual change, but do not necessitate this change to be toward a more valid construct. For example, if students become dissatisfied with the valid conception, and find a misconception to be more intelligible, plausible, and fruitful, it can result in the acceptance of the misconception, as was suggested by this study. When developing the simulation, this constructivist framework was applied with the assumption that the simulation met these critical elements for alleviating the selected misconceptions. This was not directly tested, however, and participants were not directly interviewed about these constructs. This is one limitation of the study that should be examined in future research. However, the data still suggests that the tablet-based treatments showed a greater amount of conceptual change, albeit in the opposite direction than was intended.

The current literature on seductive details theorizes that extraneous details can be detrimental to learning due to the likelihood of these details to confuse learners as to what a particular lesson is about and prime inappropriate prior knowledge about the material. (Harp & Mayer, 1998). This priming of inappropriate prior knowledge was evidenced as in the increase of the number of misconceptions expressed by participants in all conditions. Rather than leading to a sense of dissatisfaction with their misconceptions, the new information gained by participants during the treatment conditions may have been integrated with their current misconceptions, thus increasing their likelihood of expressing these misconceptions in the posttest.

One of the primary goals of this study was to investigate how the use of AR, over a non-AR, tablet-based condition, would result in a change in the number of misconceptions expressed by users of this simulation. A significant difference was discovered in the change of the number of misconceptions expressed by participants in the tablet-based, non-AR condition compared to the two AR-based conditions. A small ($d = 0.408$) effect size was shown for this effect. Participants in the tablet-based condition showed a greater increase in the total number of misconceptions expressed than participants the AR-based conditions.

It is important to note that the use of AR did not affect the change in number of expressions for all misconceptions equally. For most of the misconceptions analyzed in this study, the use of an AR-based simulation showed no significant effect over using a tablet-based simulation. However, for the misconception of intentionality, a significant difference was found. In comparison to participants in the two AR-based conditions, participants in the tablet-based conditions showed a significantly greater increase in the number of times they expressed the intentionality misconception. This change was also a larger effect size (a medium effect, $d = 0.529$) than was shown for the change in total misconceptions. This suggests that there may be additional factors influencing this effect.

While the increase in expressions of the misconception of intentionality may be partially explained by a priming of inappropriate prior knowledge, there is evidence that AR may have an effect on how participants viewed the intentionality of the entities in the simulation, and that effect resulted in a change in significant difference in expression of the intentionality misconception on the posttest.

This effect may be due to the very nature of AR. The contrast formed by the juxtaposition of computer-generated entities over the real-world emphasized the non-real nature of the AR-based entities and caused participants to find the environment less believable (Tang, Biocca, & Lim, 2004). This causes participants in the tablet-based conditions to be more likely to infer intentionality than their AR-based counterparts (Mar et al., 2007), which makes them more likely to be dissatisfied with the valid concept that the change in color exhibited by the prey entities is not due to their intentions. This dissatisfaction with the valid conception is the critical step in conceptual change that may lead participants in the tablet-based condition to show a larger increase in their expression of the misconception of intentionality.

Participants in the AR-based condition, who view the entities as non-real, are less likely to infer intentionality and thus less likely to be dissatisfied with the valid concept, and thus less likely to show an increase in their expression of the misconception of intentionality, although some may still show an increase continue to do so due to this likelihood not being reduced entirely to zero.

The second goal of this study was to analyze the effects of the type of entity in an AR simulation on the number of misconceptions expressed by participants. No significant difference was found between the symbolic and representational conditions. However, both the symbolic and representational AR-based conditions showed an increase in the number of misconceptions. While it is possible that the inability to find a significant difference between symbolic entity and representational entity conditions is due to Type II error, these results do seem to suggest that in the medium of AR the use of representational graphical entities is not more detrimental to learning over the use of symbolic entities. As hypothesized, motivation was found to not differ significantly between the two AR-based conditions. Likewise, the difference in time on task was not significant between the AR-based conditions. These results support the hypothesis that in AR-based simulations, as was seen in the literature with desktop computer simulations, the effects of a simulation are relatively equal regardless of whether the entities used are symbolic or representational.

Analyses were performed to test for an aptitude treatment interaction, and no significant results were found. This suggests that the effects found in this study apply across all levels of aptitude. Regardless of the number of preexisting misconceptions, the use of AR still results in a reduced assumption of intentionality and thus a decrease in conceptual change in regard to the misconception of intentionality. Likewise, the number of preexisting misconceptions does not appear to play a role in the amount of conceptual change caused by the use of representational and symbolic entities in AR-based simulations.

Limitations

Due to the number of participants in the study, physiological and biometric measures, qualitative researcher observations, and other methods of collecting data were not used to their full potential. These methods of data collection and analysis could have provided additional support for the results of this study. An analysis of the measures used could also provide additional supporting data. For example, it is possible that participants chose their answers based on personal assumptions on what they assumed to be the research goals of the study. This could lead them to choose answers related to predation, food gathering, coloring, and other elements from the simulation, regardless of whether those answers expressed the misconceptions chosen for this study.

Although there is no evidence that participants provided misleading information intentionally, it is possible that elements such as fatigue, boredom, or test-anxiety could have biased their answers. The pretest and posttest assessments may have been perceived as difficult and time-consuming by the participants and it is possible that they may have applied less cognitive effort to their answers on the posttest.

It was also observed that participants in the study utilized the technology in an unexpected fashion. Although none of the participants requested assistance in using the technology, many of the participants in the AR-based conditions discovered that they could stand far enough away from the environment to view the entire environment in one screen and could capture the required 75 prey entities without moving the device. This made the experience of the AR-based Representational and the non-AR conditions more similar than was anticipated and it is possible that some of the non-significant findings are the result of this increase similarity between conditions.

The design of the treatment allowed participants to take as long as they needed to capture the prey elements. This resulted in an increased time on task for participants in the AR-based conditions compared to the non-AR conditions. This opens the possibility that time on task is a confounding factor in these results. This possibility was not thoroughly examined this this study.

Lastly, the results of this study opened the possibility of additional influential constructs that were note tested directly. No direct testing on the ability of the simulation to meet the constructivist model of conceptual change was completed. It was assumed that participants would experience the elements of the model, but participants were not asked about these criteria directly. Participant views on how the simulation met, or did not meet, these constructs, particularly whether it led them to become dissatisfied with misconceptions, or even valid conceptions, was found to be lacking. Likewise, participants were not surveyed on how real or not real they viewed the simulation to be, nor were their assumptions of intentionality measured or recorded.

Future Directions

While the results of this study can point researchers in the direction of finding out why there was an increase in the number of misconceptions expressed, the results of this study alone cannot provide all the answers. One area for future research is the identification of precisely why this occurred. To do so, a more direct study on the elements required for conceptual change can be performed, particularly in regard to the element labeled dissatisfaction. While participants were not surveyed directly on the elements required for conceptual change, they were, however, given the RIMMS questionnaire, and some of the ARCS constructs that the RIMMS questionnaire is based on can be mapped to some of the constructivist approach for alleviating misconceptions. One requirement, listed as intelligible, requires that students find the new conception to be sufficiently easy to understand. It was assumed that the simulation would present the material in a simple-enough fashion that it would quality as sufficiently intelligible. Participant responses to RIMMS questions relating to their confidence in their ability to learn and use the knowledge gained from the simulation averaged a score of 4.37 out of 5, where a score of 4 represents an answer of "mostly true" and a score of 5 represents a score of "very true." Another element of the constructivist framework, plausible, requires that students find the new conception to be plausible based on their own experiences, and a third, fruitful, requires that participants be able to see how the new conception can be used to solve problems or predict phenomena. It was assumed that the simulation would meet these conditions by allowing the participants to see the valid conceptions directly implemented in the simulation and generalize those results to the real-world. Participant responses to RIMMS questions relating to the relevance of what they learned to content they already know and satisfaction that the skill is useful or beneficial to them averaged 4.01. No significant difference was found for any of the ARCS constructs on the RIMMS questionnaire. Since there was no significant difference on these constructs, it can be theorized that these constructs did not play a significant role in the conceptual change reported in this study.

Unfortunately, none of the elements of the RIMMS survey can be mapped to the final element of dissatisfaction. Testing this concept directly, by specifically asking about dissatisfaction and the nature of said dissatisfaction in future studies, could provide evidence that this element was the source of the difference in conceptual change, and further support the theory that the juxtaposition of the virtual and real worlds present in AR simulations causes AR entities to be seen as "less real" than entities in the non-AR condition, and thus less intentional.

Another possibility for studies that would extend the body of research is a detailed look into the inference of intentionality in AR-based agents and how the juxtaposition of virtual entities on the real-world can affect the assumption that AR-based agents are acting with intention.

Additionally, previous studies on the Stebbins and Allen simulation found the simulation to be valuable in alleviating misconceptions as part of a larger set of lessons and practice problems. Before completely throwing out the digital version of the simulation due to the discovered increase in misconceptions, it would be valuable to analyze it in the context of a larger series of lessons. The current design of the simulation assumed that participants would view the virtual entities as unable to have intentions. Participants were asked to think about the ability of the virtual entities to have intentions, but those priming questions may not have overcome an automatic inference of intentionality. Perhaps a separate lesson on the topic of natural selection could prevent dissatisfaction with the valid conceptions and encourage dissatisfaction with the chosen misconceptions in a way that the simulation in isolation did not.

The field of AR in education is still growing. As new technologies enter the marketplace, it is a valuable course of study to see if these results carry over to these new technologies. For example, as head-mounted displays become more commonplace, it would be valuable to recreate the simulation for use with one of these displays to see how the changes in technology affect the results. Likewise, new technologies make it easier for AR technologies to utilize the real-world environment without the use of visual markers. These technologies could allow for real-world environments, such as real rocks, plants, or other elements to be placed in the AR-based scene. Future studies could examine how using real environment objects compares to using virtual objects. As an added advantage, these environment elements could be placed in a way that require participants to move around and view the environment and prey entities from multiple angles, thus alleviating one of the limitations found in this study.

Likewise, as the technology for allowing shared AR experiences becomes more commonplace, it is valuable to look at how collaborative experiences can have an effect on the outcomes. The traditional, paper-chip version of the simulation is generally used in a collaborative space where multiple students are able to view and interact with the same prey entities and environment. This allows them to discuss what is occurring and encourages them to think critically about the material as they jointly theorize on what is happening in the simulation. This process may also help to prevent the dissatisfaction with valid theories and encourage the challenging of the associated misconceptions.

Conclusion

It has been theorized that AR can function as a powerful tool for conceptual change, such as the change required to alleviate misconceptions in complex scientific concepts such as natural selection. This study uncovered two interesting results.

First, this study did not find evidence to support concerns that the representational nature of entities used in popular AR simulations reduce AR's effectiveness as a tool for conceptual change. The results of this study do not support the theory that the use of representational entities or symbolic entities in AR simulations leads to an increase or decrease in complex misconceptions about natural selection. AR developers can feel confident that their use of representational entities is not having a significant effect on the ability of their AR simulations to drive conceptual change.

However, evidence was found that conceptual change was significantly different in the AR-based conditions over traditional, tablet-based, non-AR conditions, specifically in regard to the misconception of intentionality. While this study resulted in an unexpected increase, rather than decrease, in the number of misconceptions expressed, a significant difference was found in the number of misconceptions expressed in the AR-based conditions compared to the tablet-based conditions. This suggests that while both the AR-based and tablet-based simulations may have caused dissatisfaction with the valid conceptions, the AR-based caused less dissatisfaction, and thus less conceptual change.

This effect on the conceptual change was not equal across all misconceptions. This difference was larger in the case of one specific misconception, intentionality. A possible explanation for this may be that the juxtaposition of the real and virtual elements in an AR-based simulation reduces the impression of realism in the simulation, thus reducing the inference of intentionality, and thus resulting in less conceptual change in this one particular area. The findings of this study could be used as a basis for further research into this area, as well as inspire research into other ways in which the juxtaposition of virtual and real elements can alter perceptions of realness and influence conceptual change.

References

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

Benton, R. (1989, November). *Planned comparisons as better alternatives to ANOVA omnibus test*. Paper presented at the Annual Meeting of the Mid-South Educational Research Association, Little Rock, AR. (ED312296)

Bower, M., Howe, C., McCredie, N., Robinson, A., & Grover, D. (2014). Augmented reality in education – cases, places and potentials. *Educational Media International*, 51(1), 1-15.

Chen, P., Liu, X., Cheng, W., & Huang, R. (2017). A review of using augmented reality in education from 2011 to 2016. In *Innovations in Smart Learning* (pp. 13-18). Springer, Singapore.

Chi, M. T. (1997). Creativity: shifting across ontological categories flexibly. In T.B. Ward, S. M. Smith and J. Vaid (Eds.), *Creative Thought: An Investigation of Conceptual Structures and Processes* (pp. 209-234). Washington, DC: American Psychological Association.

Fifield, S. & Fall, B. (1992). A hands-on simulation of natural selection in an imaginary organism, platysoma apoda. *The American Biology Teacher*, 54(4), 230-235.

Ferrari, M., & Chi, M. T. (1998). The nature of naive explanations of natural selection. *International Journal of Science Education*, 20(10), 1231-1256.

Fyfe, E. R., McNeil, N. M., Son, J. Y., & Goldstone, R. L. (2014). Concreteness fading in mathematics and science instruction: A systematic review. *Educational Psychology Review*, 26(1), 9-25.

Geraedts, C. L., & Boersma, K. T. (2006). Reinventing natural selection. *International Journal of Science Education*, 28(8), 843-870.

Gregory, T. R. (2009). Understanding natural selection: Essential concepts and common misconceptions. *Evolution: Education and Outreach*, 2(2), 156-175.

Gregory, T. R., & Ellis, C. A. (2009). Conceptions of evolution among science graduate students. *BioScience*, 59(9), 792-799.

Hewson, M.G., & Hewson, P.W. (2003). Effect of instruction using students' prior knowledge and conceptual change strategies on science learning. *Journal of Research in Science Teaching*. 40(S1), S86-S98.

Harp, S. F., & Mayer, R. E. (1998). How seductive details do their damage: A theory of cognitive interest in science learning. *Journal of Educational Psychology*, 90(3), 414-434.

Johnson, L. F., Levine, A., Smith, R. S., & Haywood, K. (2010). Key emerging technologies for elementary and secondary education. *Education Digest*, 76(1), 36–40.

Kaminski, J., Sloutski, V., & Heckler, A. (2008). The advantage of abstract examples in learning math. *Science*, 320, 454-455.

Keller, J. M. (2010). *Motivational design for learning and performance: The ARCS model approach*. New York, NY: Springer

Kwon, M. (1996, January). *The use of planned comparisons in analysis of variance research*. Paper presented at the Annual Meeting of the Southwest Educational Research Association, New Orleans, LA.

Loorbach, N., Peters, O., Karreman, J., & Steehouder, M. (2015). Validation of the Instructional Materials Motivation Survey (IMMS) in a self-directed instructional setting aimed at working with technology. *British journal of educational technology*, 46(1), 204-218.

Mar, R. A., Kelley, W. M., Heatherton, T. F., & Macrae, C. N. (2007). Detecting agency from the biological motion of veridical vs animated agents. *Social cognitive and affective neuroscience*, 2(3), 199-205.

Mayer, R. E. (Ed.). (2005). *The Cambridge handbook of multimedia learning*. New York, NY: Cambridge University Press.

Meagher, T. R. (1999). Evolution and today's society. *BioScience*, 49(11), 923-925.

National Academy of Sciences (US). Working Group on Teaching Evolution. (1998). *Teaching about Evolution and the Nature of Science*. Washington, DC: National Academy Press.

Paneta, K. (2017, August 15). *Top Trends in the Gartner Hype Cycle for Emerging Technologies, 2017*. Retrieved from https://www.gartner.com/smarterwithgartner/top-trends-in-the-gartner-hype-cycle-for-emerging-technologies-2017/

Perkins, D. N., & Simmons, R. (1988). Patterns of misunderstanding: An integrative model for science, math, and programming. *Review of Educational Research*, 58(3), 303-326.

Pintrich, P. R. (2003). A motivational science perspective on the role of student motivation in learning and teaching contexts. *Journal of Educational Psychology*, 95(4), 667.

Plass, J. L., Homer, B. D., Milne, C., Jordan, T., Kalyuga, S., Kim, M., & Lee, H. (2009). Design factors for effective science simulations: Representation of information. *International Journal of Gaming and Computer-Mediated Simulations (IJGCMS)*, 1(1), 16-35.

Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in cognitive sciences*, 4(8), 299-309.

Sinatra, G. M., Brem, S. K., & Evans, E. M. (2008). Changing minds? Implications of conceptual change for teaching and learning about biological evolution. *Evolution: Education and outreach*, 1(2), 189-195.

Stebbins, R. C., & Allen, B. (1975). Simulating evolution. *The American Biology Teacher*, 37(4), 206-211.

Tang, A., Biocca, F., & Lim, L. (2004). Comparing differences in presence during social interaction in augmented reality versus virtual reality environments: An exploratory study. *Proceedings of PRESENCE*, 204-208.

Windschitl, M. A. (1995). Using computer simulations to enhance conceptual change: The roles of constructivist instruction and student epistemological beliefs. *Journal of research in science teaching,* 35(2): 145–160.

Yoon, S., Wang, J. (2014). Making the invisible visible in science museums through augmented reality devices. *TechTrends*, 58(1), 49-55

Yoon, S., Anderson, E., Lin, J., & Elinich, K. (2017). How augmented reality enables conceptual understanding of challenging science content. *Journal of Educational Technology & Society*, 20(1), 156.

Zvinakis, A. (2017, November 1). Poly: Browse, discover and download 3D objects and scenes [Blog Post]. Retrieved from https://www.blog.google/products/google-vr/poly-browse-discover-and-download-3d-objects-and-scenes/

APPENDIX A

PRESURVEY

Presurvey

ID: ____

Gender: _____________ Age: ___________ Race/Ethnicity: ________________

Major/Degree: ______________ College Semesters Completed: ________________

How would you rate your familiarity with Computer-based Education / Learning?
Very Familiar Familiar Neither Familiar or Not Familiar Not Familiar

How would you rate your familiarity with Augmented Reality and Mixed Reality?
Very Familiar Familiar Neither Familiar or Not Familiar Not Familiar

How would you rate your education level in biology?
Excellent Good Fair Poor

How would you rate your attitude towards biology?
Very Positive Positive Neutral Negative Very Negative

How would you rate your familiarity with natural selection?
Very Familiar Familiar Neither Familiar or Not Familiar Not Familiar

How would you rate your attitude towards natural selection?
Very Positive Positive Neutral Negative Very Negative

How would you rate your familiarity with mobile devices, such as tablets and smartphones?
Very Familiar Familiar Neither Familiar or Not Familiar Not Familiar

How would you rate your ability to use mobile devices, such as tablets and smartphones?
Excellent Good Fair Poor

How would you rate your attitude towards mobile devices, such as tablets and smartphones?
Very Positive Positive Neutral Negative Very Negative

APPENDIX B:

POST-ACTIVITY SURVEY

Post-Activity Survey

There are 12 statements in this questionnaire. Please think about each statement in relation to the instructional materials you have just studied and indicate how true it is. Give the answer that truly applies to you, and not what you would like to be true, or what you think others want to hear.

Think about each statement by itself and indicate how true it is. Do not be influenced by your answers to other statements.

	not true	slightly true	moderately true	mostly true	very true
1. The quality of the material helped to hold my attention.					
2. The way the information was arranged helped keep my attention.					
3. The variety of information, helped keep my attention on the lesson.					
4. It is clear to me how the content of this simulation is related to things I already know.					
5. The content and style in this lesson convey the impression that being able to use the information presented is worth it.					
6. The content of this simulation will be useful to me.					
7. As I worked with this simulation, I was confident that I could learn the information presented.					
8. After working with this simulation for a while, I was confident that I would be able to complete exercises using the information presented.					
9. The good organization of the content helped me be confident that I would learn to work with the information presented.					
10. I enjoyed working with this simulation so much that I was stimulated to keep on working.					
11. I really enjoyed working with this simulation.					
12. It was a pleasure to work with such a well-designed simulation.					

APPENDIX C:

PRETEST / POSTTEST

Pre/Post Test

Participant ID#________________________

Your answers will be confidential. In every case below, "evolution" means "biological evolution".

1. How would you describe the process of natural selection?

2. In the lowest parts of the Pacific Ocean lives a species of fish with thin, translucent scales. The closest known ancestor species lives in shallower waters and has much thicker and darker-colored scales. Using your knowledge of natural selection, explain why you think these fish have thin scales.

3. In the area surrounding an active volcano, a new species of bird was recently discovered that is brightly colored in comparison to its nearest genetic relative. Using your knowledge of natural selection, explain why you think this species of bird is brightly colored.

Instructions: Read the following scenarios. For each scenario, indicate ALL of the statements that are likely to be TRUE, in terms of how evolutionary biologists use and understand the theory of evolution today. You can choose more than one answer.

4. A species of mouse native to the rivers of Thailand was introduced to nearby city sewers fifty years ago. Just yesterday, some of these mice were captured and compared to rats that still live in the rivers and marshes. Select all of the statements that are likely to be TRUE, in terms of how evolutionary biologists use and understand the theory of evolution today.

> a. The mice in the sewers need to develop different features from the mice in the rivers and so the two species will be very different.
>
> b. The mice in the sewers will have sought out mates with genetic traits that are better suited for the urban environment and will have evolved into a different species.
>
> c. The mice in the sewers will have access to high-fat processed foods and will have evolved to be fatter than their river-dwelling relatives.
>
> d. The two sets of mice will be very similar because the mice would not yet have started to evolve.
>
> e. It is possible that the mice could have evolved to be similar, or it is possible that they could have evolved to be significantly different. There is not enough information in this statement to say how similar or different the two rats will be.

5. The plains of southern Africa are home to two of the world's fastest creatures, the gazelle and the cheetah. Select all of the statements that are likely to be TRUE, in terms of how evolutionary biologists use and understand the theory of evolution today.

> a. In order to provide their offspring with a better chance of survival, female gazelles choose mates who are able to outrun cheetahs.
>
> b. Cheetahs needed to become faster in order to catch and eat gazelles. Likewise, gazelles needed to become faster in order to avoid being eaten. Therefore, the two species evolved to meet those needs.
>
> c. Individual gazelles developed mutations that allowed them to better outrun predators. These survivors passed these genes on to their offspring. Likewise, individual cheetahs also passed mutations on to their own offspring.
>
> d. Chasing after prey caused adult cheetahs to develop larger leg and chest muscles. They passed these traits on to their offspring. Meanwhile, adult gazelles

who developed similar traits over their lifetime passed their traits on to their own offspring.

e. The two species evolve together as a shared system. If the gazelle population were to be removed from the plains, the cheetahs of Africa would stop evolving.

6. Two species of bears are very similar in features, except for the thickness of a subdermal fat layer located directly under their fur. One species of bear lives in a colder climate. Select all of the statements that are likely to be TRUE, in terms of how evolutionary biologists use and understand the theory of evolution today.

a. Genetic variation causes different bears in each species to have different amounts of fat in their subdermal layer.

b. Because plant-based foods are harder to come by in colder climates, bears in the colder climates consumed more fat calories and stored those calories as fat. They passed this weight gain on to their offspring.

c. Bears in the colder climates became uncomfortable in the cold and evolved an extra layer of fat to feel warmer.

d. If the bears without the subdermal layer of fat were to move to a colder climate, they would begin to evolve.

e. The bears that lived in the colder client developed an extra layer of fat because they need this fat to survive the cold environment.

7. Hunters have noticed that the antlers of male deer (bucks) in the woods of Northern Michigan grow at a slower rate than noted in previous years. Data also reports an increase in the amount of hunting in the area, and that bucks with larger antlers are more likely to be taken by hunters. Select all of the statements that are likely to be TRUE, in terms of how evolutionary biologists use and understand the theory of evolution today.

a. When hunters began to increase the amount of hunting in the area, it triggered the local bucks to start the process of evolution and adaptation.

b. Increase hunting caused bucks to be more active, which resulted in additional muscle growth. This muscle growth prevented the growth of large antlers and the bucks passed this reduced antler growth on to the next generation.

c. The bucks evolved to grow their antlers at a slower rate so that they would not be as likely to be killed by hunters seeking large-antlered bucks.

d. Some of the bucks outsmarted the hunters by adapting to have slower-growing antlers.

e. Some bucks had genetic traits that allowed them to survive hunting season and reproduce. They produced offspring who also carried those traits.

8. Individuals in a species of butterflies show a wide variation in wing size and shape. Researchers in the area have discovered a much larger number of individuals with more complex, bright-colored wings than those with simpler, dark-colored wings. Select all of the statements that are likely to be TRUE, in terms of how evolutionary biologists use and understand the theory of evolution today.

a. Once enough of these butterflies are able to successfully find mates, they will stop evolving and become a new species.

b. Some of the traits expressed by these butterflies are due to genetic diversity within the species.

c. Individuals with the simpler wing shape must have evolved first, and the more complex wing shapes later evolved to better fit the environment.

d. As the number of brightly colored butterflies in the area increases, many predators will decide to evolve better visual perception to help them discern between species.

e. Since more complex wing shapes are more likely to be damaged, it is likely that these individuals will damage their wings and pass that damaged wing type on to their offspring.

9. The first time Susan used an antibacterial spray on her garbage can, the number of bacteria on the surface of the can decreased significantly. However, after a few weeks of applying the spray, the number of bacteria that survived each spray began to increase.

a. The bacteria evolved to become immune to the spray because they needed to in order to survive.

b. When Susan began to spray the garbage can, it triggered the bacteria to start the process of evolution and adaptation.

c. The bacteria outsmarted Susan and chose to adapt in order to survive despite her spraying.

d. Exposure to the spray caused some of the bacteria to build a thick, protective cell wall. They passed this thick wall on to the next generation.

e. Some bacteria had genetic traits that allowed them to survive the initial application of the antibacterial spray. They produced offspring who also carried those traits.

10. A team of biologists notices a complex predator/prey relationship between a species of owls and a species of mice in a field in the south-western United States. Over multiple generations, the mice have adapted the color of their coats to blend in with the soil and the owls have adapted their eyesight to better see small, fast-moving objects.

a. The two species evolve together as a shared system. If the owl population were to be removed from the plains, the mice would stop evolving.

b. To better see the mice they were hunting, several individual owls needed to hold their eyes open larger. This led them to developing larger eyes and they passed these traits on to their offspring.

c. In order to provide their offspring with a better chance of survival, female owls choose mates who are able to see smaller objects.

d. Individual owls developed random mutations, some of which allowed them to better see their prey. Likewise, individual mice also developed random mutations, some of which allowed them to better hide in the dirt.

e. Owls needed to develop better vision in order to catch and eat mice. Likewise, mice needed to blend into their surroundings in order to avoid being eaten. Therefore, the two species had to evolve to meet those needs.

(Posttest Only)
11. You recently used a digital simulation on natural selection where you acted as a predator capturing prey. At the end of the simulation, you were asked if you saw a pattern, and what you think occurred. Did you see a pattern? Why do you think that pattern occurred?