WHAT'S CLOSENESS GOT TO DO WITH IT? INVESTIGATING THE EFFECTS OF INTERFACE CLOSENESS ON ABSTRACT PROBLEM SOLVING AND LEARNING

Thomas James Donahue

A Thesis

Submitted to the Graduate College of Bowling Green State University in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

August 2012

Committee:

Dr. G. Michael Poor, Advisor
Dr. Laura Leventhal
Dr. Guy Zimmerman
Dr. Dale Klopfer
This work investigated the effects of interface closeness on abstract problem solving and learning. Interface closeness is the degree to which an action performed via the interface differs from the action performed to achieve the desired result in the natural world. Interface closeness has numerous theoretical problem solving and learning benefits, though empirical investigations have found weak and conflicting results. A comparative study between interfaces at three levels of closeness (mouse, touchscreen, and tangible) on a novel abstract problem solving task was conducted.

We found that the tangible interface was significantly slower than both the mouse and touchscreen interfaces. However, the touchscreen and tangible interfaces were shown to be significantly more efficient than the mouse interface in problem solving across a number of measures. Overall, it was shown that the touchscreen condition was the best interface, but also that closer interfaces in general offer significant benefit over the traditional mouse interface on abstract problem solving and learning.
Tada gan iarracht
ACKNOWLEDGMENTS

I would first like to thank my committee for guiding me through this process. Throughout my academic career, they have always been there to share their wisdom, their time and a kind word. Additionally, I would be remiss if I did not thank Dr. J. Devin McAuley, who instilled in me the traits for success as a researcher and mentor. I would also like to thank my lab mates, both past and present, for sharing their time and thoughts in our windowless cave. In particular, I would like to extend a special thank you to Martez Mott—an irreplaceable friend and colleague.

Furthermore, I would like to thank my family, especially my mother and father, who have given more than I can ever express. Last, but certainly not least, are my wife Lindsay and daughter Alex. Though the journey has been long and arduous, they have always been by my side. Thank you for your patience and your love.

Sláinte
Table of Contents

CHAPTER 1: INTRODUCTION ................................................................. 1

1.1 Closeness .................................................................................. 2

1.1.1 Tangible User Interfaces ....................................................... 6

1.2 Discovery Learning ................................................................. 8

1.3 Purpose .................................................................................. 9

1.4 Organization ........................................................................ 10

CHAPTER 2: RELATED WORK .......................................................... 11

2.1 Tangible Interfaces for Learning ............................................... 11

2.1.1 Learning Domains ............................................................. 11

2.1.2 Possible Learning Benefits ................................................ 13

2.1.3 Comparison of Tangible versus Traditional ...................... 16

2.1.4 Discussion ........................................................................ 19

2.2 Intersection of Interface and Problem Solving ..................... 21

2.2.1 Discussion ....................................................................... 24
CHAPTER 3: TASK AND SYSTEM DESCRIPTION ............................................... 26

3.1 Task ........................................................................................................... 26

3.1.1. Task Description .................................................................................. 27

3.1.2. Arrangement Creation ......................................................................... 30

3.1.3. Task Benefits ...................................................................................... 31

3.2 Software ..................................................................................................... 32

3.3 Physical Setup ........................................................................................... 33

3.3.1. Tangible Condition ............................................................................ 34

CHAPTER 4: METHODS ..................................................................................... 36

4.1 Introduction and Hypotheses ..................................................................... 36

4.2 Experimental Design .................................................................................. 37

4.3 Participants ............................................................................................... 37

4.4 Experimenters ........................................................................................... 37

4.5 Procedure .................................................................................................. 38

4.6 Data Collection .......................................................................................... 39

4.6.1. Demographics .................................................................................... 39

4.6.2. Training and Experimental Trials ....................................................... 40

4.6.3. TLX Questionnaire ............................................................................ 40

CHAPTER 5: RESULTS ....................................................................................... 41

5.1 Analysis of Performance .......................................................................... 41

5.1.1. Speed ................................................................................................. 42
List of Figures

1.1 Norman’s Seven Stages of Action ................................................................. 3

1.2 Artistic rendering of the Marble Answering Machine ................................. 7

3.1 Example of trial arrangement from experimental task &

   Example solution path taken by a participant .................................................. 29

3.2 Screenshot of Mouse/Touch software .......................................................... 32

3.3 Setup of the tangible condition .................................................................... 34

5.1 Average Time Taken per Trial ...................................................................... 41

5.2 Average Number of Blocks per Trial ............................................................ 42

5.3 Average Submissions per Trial ..................................................................... 43

5.4 Average Time Taken per Submission ............................................................ 43

5.5 Summary of results from the NASA TLX .................................................... 44
CHAPTER 1

INTRODUCTION

Computing systems are being integrated into a growing number of areas in our lives. When we come in contact with these systems in obvious contexts, such as our homes, offices, or even the grocery store, our relationship with them and the role they play is well defined. This is to say, in most cases, these systems are used as a tool we employ to get something done. However, as these systems expand and integrate into less obvious contexts, such as elementary school classrooms, they become more than just tools to get something done, but actually augment the learning experience for the better. As such systems become more pervasive in our lives, the very nature of the interaction plays an enormous role in shaping the experience. Different settings have different goals and constraints which require different outcomes from the interaction. Systems in the classroom environment, for example, would require an interface that is both engaging and fosters learning. As a result, practitioners looking into such domains should utilize interaction systems which promote these factors. One element that is thought to yield such benefits is the closeness of the interface.
1.1 Closeness

The psychological basis to the concept of closeness can be traced to J.J. Gibson’s seminal visual perception work in ecological psychology [10]. Central to Gibson’s theory of visual perception is the idea of *affordances*. Affordances can be thought of as possibilities for action, which exist in the environment but are in relation to the actor present in the environment. Thus, a pencil would not afford the act of writing if, for example, the actor had severe arthritis in their hands. According to Gibson, the environment is perceived, at least in part, in terms of what it affords an organism.

In this same vein, but in a more applied context, is the work of Don Norman. In *The Design of Everyday Things* [32], he restricted Gibson’s idea of affordances and introduced the “seven stages of action.” Norman restricted affordances to those which are easily perceivable by the actor in the environment. This restriction is important in that an affordance depends on more than just the physical ability of the actor to interact with it, but that the actor’s goals in an environment are relevant as well. This actor-centric approach is also present in Norman’s seven stages of action (see Figure 1.1).
Figure 1.1: Norman’s Seven Stages of Action [32]

The seven stages of action are Norman’s way of describing the psychology behind an actor’s performance of a task. The seven stages can be broken down into two distinct sub-stages: execution and evaluation. The stages of execution are: set the goal, translate the goal into intended actions, formulate an action sequence to carry out the intended actions, and carrying out the action sequence. The stages of evaluation are: perceive the world, interpret our perception in context of our expectations, and analyze our interpretations in reference to our intentions and goals. Coupled with these stages are two corresponding gulfs. The gulf of execution is the disparity between the intended actions and the affordances, while the gulf of evaluation refers to how difficult it is to determine if our intentions and expectations line up with our interpretation of the system. To illustrate the gulf of execution, and its relationship to closeness, take the relatively simple computing task of deleting a file from the desktop in both a mouse and touchscreen interaction. This task can be, and often is, accomplished by dragging the file to the Recycle Bin. From a user’s point of view, the perceived
actions to accomplish this task would simply be ‘drag the file to the Recycle Bin.’ However, in reality, in order to accomplish this task, the required actions are more detailed and complex. In a mouse interaction users must move the cursor to the file, click and hold the left mouse button on the file they wish to delete, drag the file to the recycle bin, then release the left mouse button. In contrast, for the touchscreen interaction users must touch the file, drag it to the recycle bin, and then release their finger. The difference between users’ perceived actions and the required actions is smaller for the touchscreen interaction, thus the gulf of execution is smaller. The sizes of the gulfs inversely correlate with the closeness of the interaction (i.e. the larger the gulfs of execution and evaluation are, the less close the interaction is).

Extending the work of Gibson in ecological psychology, Vincente and Rasmussen proposed an interface design framework known as ecological interface design (EID) [42]. The goal of EID is to shift a user’s cognitive workload away from the lower level elements of the interface and toward higher level understanding of the goal to be accomplished. In this regard, an interface which achieves the EID goal would be closer, as the gulf of execution and evaluation should be smaller for such an interface.

Though there is a strong psychological and HCI basis for closeness in interfaces, integration of closeness into a modern framework has only been a recent development. The Reality-Based Interaction (RBI) framework [20] sought to describe, and coalesce, post-WIMP interactions (i.e. traditional Windows, Icons, Menus and Pointing device interactions), in terms of user knowledge, under a new framework based on various themes of reality. There are four themes which comprise
the framework: naïve physics, body awareness and skills, environmental awareness and skills and social awareness and skills. The first three are of importance to the concept of closeness. Naïve physics (NP) is “the informal human perception of basic physical principles,” which includes common sense knowledge of concepts such as gravity, friction and velocity. For example, a tangible user interface (TUI; see Section 1.1.1) which allowed players to interact with a computer chess game using a physical chess board and pieces. In relation to closeness, NP plays a vital role in that, if an interface were inconsistent with our common sense knowledge, cognitive resources would have to be devoted to dealing with such inconsistencies, instead of focusing on the task to be accomplished. Body awareness and skills (BAS) is the “familiarity and understanding that people have of their own bodies, independent of the environment.” As an example, special treadmills are often utilized in virtual reality (VR) interactions to allow users to “walk” about the virtual environment. Like NP, BAS’s role in closeness is best illustrated by the detrimental effect of its absence. That is, if an interface were to ignore a user’s implicit knowledge of their body and its movements, the user would once again need to devote cognitive resources to adapting to this interface in an unnatural manner. Finally, environmental awareness and skills (EAS) is the sense people have of their surroundings, and how to negotiate within them. For instance, in many VR and tangible interfaces, users often select interface elements by physically or virtually grasping it. In terms of closeness, an interface which takes into account and leverages what the user implicitly knows about their surroundings will place the interaction within a more natural context.

The concept of closeness, as used in this thesis, was first introduced by Hippler et. al in More than Speed? An Empirical study of Touchscreens and Body Awareness on an Object Manipulation Task.
In the article they synthesize the work of [10, 42] to create a lucid, foundationally strong definition of closeness. With that said, they acknowledge that their current definition is not yet sufficient to yield predictive value [15].

1.1.1 Tangible User Interfaces

One particularly relevant class of interfaces to the discussion of closeness are tangible user interfaces. Though aspects of TUIs had been around for a number of years (e.g. [9]), it wasn’t until 1997 that they were concretely defined. In Ishii and Ullmer’s *Tangible Bits* paper, they describe TUIs as any user interface that “augments the real physical world by coupling digital information to everyday physical objects and environments [19].” Thus, by definition, tangible interfaces are inherently close. The motivation behind the introduction of these interfaces was the “rapid flood of digital technologies” that replaced the rich “historical scientific instruments, most of which have disappeared from schools [and] laboratories.”

The range of domains in which TUIs have been utilized is vast. An early example, which predates the conception of the term, was the Marble Answering Machine, created by Durrell Bishop (see Figure 1.2) [37].
Figure 1.2: Artistic rendering of the Marble Answering Machine [37]

Just as any other answering machine, the system would store and playback messages as well as allow the user to call back those whose calls they had missed. However, unlike traditional answering machines, the messages were not stored on tape or digitally, but rather linked to a physical marble. In order to playback the message, the user had to pick up the marble and place it in a specific indentation on the machine. Similarly, to redial the person whose call they had missed, the user had to pick up the marble and place it in a different indentation on the machine. Though a precursor to the Ishii and Ullmer’s *Tangible Bits*, it is a perfect example of an interface that “augments the real physical world by coupling digital information to everyday physical objects and environments.” Other examples, from the tangible video editor [44] to the tangible programming [17] and password [29] systems illustrate the versatility of tangible interfaces.
1.2 Discovery Learning

Discovery learning, sometimes referred to as exploratory learning, is based on the constructivist theory of learning, which can be traced to John Dewey. The basic premise of constructivism is that knowledge and understanding is actively constructed by the learner. Discovery learning, therefore, can be thought of as the process by which the active construction takes place. Though there is no clear originator to the idea of discovery learning, its modern development is often credited to Jerome Bruner, with Jean Piaget, John Dewey, and Seymour Papert being significant contributors. More recently, Wouter van Joolingen in *Cognitive tools for discovery learning* [41] described it this way:

Discovery learning is a type of learning where learners construct their own knowledge by experimenting with a domain, and inferring rules from the results of these experiments. The basic idea of this kind of learning is that because learners can design their own experiments in the domain and infer the rules of the domain themselves they are actually constructing their knowledge. Because of these constructive activities, it is assumed they will understand the domain at a higher level than when the necessary information is just presented by a teacher or an expository learning environment.

Though not the focus of this thesis, there is on-going debate in the research community as to the efficacy of discovery learning as opposed to other approaches, such as guided and problem-based learning. For further information about the arguments against discovery learning, see [21, 26].
1.3 Purpose

As the role of computer systems change and become more integrated into our society, the way that we interact becomes increasingly important. One of the most important roles these new systems play is that of augmenting learning. In such systems, the nature of the interaction plays an essential role in shaping the experience, but traditional interaction styles are less effective in this context.

Practitioners have looked to closer interaction styles, specifically TUIs, as a solution to this problem. While the conventional wisdom is that TUIs should yield a better, more robust learning experience, emerging research has found conflicting results. With the efficacy of TUIs for learning called into question [24], it is important for the research community to step back and build up the foundational knowledge as to what features, characteristics and style of interaction yield the best outcome for learning, whether it is tangible or something else entirely.

This thesis presents an investigation into the effects of interface closeness on abstract problem solving and learning. The contribution of this investigation is multifaceted. First, it will offer some resolution to the debate as to whether interfaces which are closer, such as touch and, in particular, TUIs, offer significant benefit over a more traditional interface/interaction (i.e. mouse) for learning. Second, the task explores abstract problem solving and learning, a domain where TUIs are not typically thought to be of much benefit. Though the task leverages aspects of discovery learning, it itself is not a strict discovery learning task. This investigation will empirically test such assumptions. Finally, if closer interfaces do offer significant benefit, this investigation compare and
contrast the learning benefits of an interaction which is simply closer (i.e. mouse vs. touch) against those of a closer interaction which also has physicality (i.e. mouse/touch vs. tangible).

1.4 Organization

The remainder of this paper is organized as follows: Chapter 2 will give an overview of related work; Chapter 3 will provide a description of the system and the task present in the study; Chapter 4 will describe the methodology of the user study; Chapter 5 will present the results of the study; and Chapter 6 will include a discussion of the results, limitations, plans for future work and conclusions.
CHAPTER 2

RELATED WORK

2.1 Tangible Interfaces for Learning

As TUIs grow in popularity, the need to develop and test new TUIs to discover their benefits over more traditional systems becomes increasingly necessary. Unfortunately, much TUI research has relied on intuition and assumption, rather than empirically supported psychological and cognitive theories [4]. This same trend was seen, and roundly criticized by many, in early graphical user interface (GUI) work [3, 13], as such assumptions were often disproved through later empirical investigation. In contrast, the domain focus of this thesis—tangible interfaces for learning—is built on top of a stronger psychological and educational research foundation. This section will discuss the learning domains most commonly associated with tangible interfaces, introduce and explore the possible learning benefits afforded to tangible interfaces, as well as describe and analyze a number of studies which have directly compared TUIs to GUIs.

2.1.1 Learning Domains

Tangible interfaces, in general, have been applied to a number of diverse domains and problems (e.g. [17, 29, 44]). On the surface, TUIs for learning appear to have been utilized in a similarly diverse number of applications. Upon closer inspection, however, there is a common element present in
much of this apparent diversity. This common element, the inherently spatial structure, in a physical or metaphorical sense, of the domains/problems is manifested in diverse ways across applications [24].

Two TUI implementations that illustrate the physical and metaphorical spatiality for which TUI are most commonly used for are Gillet et. al’s *Tangible Interfaces for Structural Molecular Biology* [11] and Horn and Jacob’s *Designing Tangible Programming Languages for Classroom Use* [17]. In the former, researchers utilized a 3D printer to create molecular models which varied greatly in size, shape and complexity. Utilizing augmented reality, which is a combination of a real world object/scene and virtual data, they attached tracking codes to different parts of the molecules and tracked their position and orientation in real time. As they manipulated the physical molecular model, on a computer screen the system overlaid a virtual version of the molecule with extra information about the molecule, such as electrostatic fields and electron clouds. Preliminary results found the augmented molecules to be “quite engaging and instructive.”

In *Designing Tangible Programming Languages for Classroom Use*, Horn and Jacob created two programming languages: Quetzal and Tern. Quetzal was used to control Lego Mindstorms robots and consists of interlocking physical tiles which “represent flow-of-control structures, actions, and parameters.” These physical tiles are then linked together to create simple commands and basic programs. Similar to Quetzal, Tern is composed of interlocking physical pieces, but instead of controlling Lego Mindstorms robots, it controls virtual robots. Both Quetzal and Tern are implemented by having each of the physical tiles augmented with a circular, barcode-like symbol
known as a SpotCode. SpotCodes allows the system to quickly, via a camera, pick up the size, position and type of command attached to each physical tile. Initial results with a group of first and second grade students were positive. Students were “easily able to construct flow-of-control chains”, collaborate, and even find bugs in their code.

As some researchers have pointed out [24], however, there is no strong empirical basis for the focus on learning tasks which are inherently spatial, in either sense. In fact, there is no strong empirical basis to support the notion that TUIs offer significant learning benefits over their traditional GUI counterpart in such spatial tasks. Such benefits are supported theoretically, but results from empirical investigations are conflicting (see Section 2.1.3).

2.1.2 Possible Learning Benefits

Tangible interfaces offer, by their very nature (i.e. physical, hand-centric manipulation), certain qualities that other interfaces cannot. The most important of these qualities, in regard to learning benefits, are the physicality and concreteness of TUIs. Though physicality and concreteness are similar, in that they both deal with the role physical manipulatives play in shaping our cognition, the distinction between the two ideas is important [5].

Physicality can be conceived of as the effect that the physical nature of an object plays during its manipulation on the cognitive processes of the manipulator. The role of physicality on our cognitive processes is based on the cognitive science work of embodied cognition. Embodied cognition’s key tenet is that our bodies (i.e. our motor system and the sensations we take in) influence the way we think and process information, with the opposite being true as well. Strack,
Martin and Stepper’s 1988 experiment [38], one of the most well-known and replicated [2, 35], illustrates this tenet quite clearly. Participants were to hold a pen in their mouth with their teeth, or with their lips, which forced a partial smile or partial frown, respectively. Following some auxiliary tasks, participants were asked to rate the funniness of cartoons (of the magazine variety) on a scale of 0 to 9, with 9 being very funny. The results showed that participants that held the pen in their lips rated the cartoons consistently less funny than those who held the pen in their teeth, supporting the idea that our bodies affect our cognitions. Relating this research to physicality and TUIs, given the strong link between our bodies and our cognition [1], as well as the increasing body of research that suggests higher order thinking is built from, and out of, our sensory motor system [23], the act of physical manipulation with TUIs may offer benefits over traditional GUIs (see Section 2.1.3). Furthermore, moving away from cognitive science, Maria Montessori’s educational approach, deemed *The Montessori Method* [28], adopts and extends the work of constructivism and discovery learning (see Section 1.2). As a result, it stresses the importance of manipulating physical materials in order to foster learning.

In contrast to physicality, concreteness can be thought of as the way concepts, or knowledge, are tied to physical representations and their manipulations [27]. The notion of concreteness can be traced to Piaget’s developmental psychological work. An example of this would be the way young children learn to add two numbers together, such as 2+3. Often, children are taught to link the abstract concepts of the numbers 2 and 3 to physical objects, such as their fingers or apples, and then count out the physical objects in order to compute the sum. Applying this notion to TUIs, because TUIs, by Ishii and Ullmer’s [19] definition act as a bridge between the physical world and the
abstract digital world, thus TUIs may aid in the linking of an abstract concept to a concrete notion in a user’s mind. Revisiting the tangible programming work of Horn and Jacob [17], the abstract computer programming notions, such as a loop or a branch, are made concrete through the physical manipulatives of the tangible programming system.

Though concrete manipulatives are traditionally thought of as physical, it should be noted that some researchers have explored the intersection of concreteness with the virtual world. Research into virtual manipulatives (i.e. computer-based interactive systems which attempt to mimic the real world richness of physical manipulatives) has shown promise, especially in mathematics education [30]. Though virtual manipulatives attempt to mimic the richness of physical objects, they often implement idealized (i.e. abstracted, less detailed) versions instead. Traditionally, it has been thought that the more realistic the manipulative is, the better it is for use in learning materials. This notion, however, has recently been called into question by a number of researchers. For example [12] Goldstone and Son claim that idealized manipulatives offer a number of advantages over their rich, detailed counterparts. They claim that knowledge gained through interactions with idealized manipulatives may be more transferable across domains, and that they may speed up the transfer from manipulatives to symbolic mental representations. Further, empirical work by DeLoache [6, 7], explored the use of physical models as representations in young children (roughly 2 1/2 years old). In the study, children were given a model of a room, shown a miniature toy hidden in the model room and then told that a large version of said toy was hidden in the same place in the real room. It was shown that the children were better at finding the toy in the real room when the model was a two-dimensional picture, rather than a physical three-dimensional replica.
2.1.3 Comparison of Tangible versus Traditional

For many years, the theoretical benefits of TUIs for learning from various fields (e.g. psychology, cognitive science, neuroscience and education), had gone untested. In recent years, however, a number of investigations have been carried out to test the predicted theoretical benefits of TUIs by conducting comparison studies between TUIs and the more traditional GUI.

One of the first to conduct a direct comparison study was Fails et al.’s Childs Play [8]. The comparison was carried out between tangible and graphical versions of the Hazard Room Game. The Hazard Room Game, which was created for the study, is a game with the goal of teaching children about environmental health hazards. Both quantitative and qualitative data were collected for the study. The quantitative data consisted of pre- and post-tests that were given to each participant, which consisted of a number of questions about environmental hazards (e.g. “What should you do with an apple before you eat it?”). Qualitative data consisted of notes, and video taken during discussions between the children and researchers which were then coded based on a number of criteria. No statistical differences were found between the tangible and graphical conditions in learning about the environmental hazards. However, when the descriptive statistics were compiled in conjunction with the qualitative data, they painted a different picture. Participants in the tangible condition were more interested, engaged, and had increased depth of response when discussing their understanding of the environment.

While the other comparison studies had explored physicality to some degree, by the nature of comparison, none has adequately controlled for the covariates with physicality [40]. The
following two studies were carried out to explicitly explore the theoretical benefit of physicality. The first of these two studies *Hands on What?* [22] focused on seventh- and eighth-grade student participants designing and testing physical and virtual mousetrap cars. Mousetrap cars are, as their name implies, small wooden cars which are powered by a mouse trap. The challenge for the participants in the study was to discover the set of features that would produce a car that travelled the farthest. In designing their car, there are four features that could be changed: body type, back axle, back wheels, and front wheels. Participants were divided into groups by build material and build constraint. Build material dealt with what they would be building their cars out of, while build constraint deals with the way they would carry out the build task. Build material had two levels: virtual or physical. Similarly, build constraint had two levels: fixed amount of time to build in and fixed amount of cars to build. Upon completion of the experiment, a post-test which probed the participant’s knowledge of the task, which included the participant having to build the car they felt would travel the farthest. No significant differences were found in participant’s knowledge by build material condition or by build constraint condition. Participants in the virtual build condition did, however, accomplish the task significantly faster.

The second study, *Tangibles in the Balance* [25], investigated adults’ discovery learning on a balance beam task. The basics of the balance beam task, in its common form, ask the participant to predict and explain the movement of the beam based on the weights that are placed along it. The predictions for the movement of the beam in response to the weights can be one of three options: left side down, right side down or keeping balance. Participants were divided into groups based on two conditions, beam materials and experiment control. The beam materials condition had two
levels: physical or virtual. The experiment control condition also had two levels: with control and without control. In the with-control condition, participants carried out experiments on the balance beam of their own design, while those in the without-control condition carried out experiments that were given to them. Two types of tests were given to participants to measure the participants’ level of understanding: motor performance and written tests. The motor performance test measured participants’ accuracy in positioning weights in order to balance out the beam, while the written tests measured the participants’ understanding of higher order concepts, such as balance. On the motor performance test, no significant main effects by the material condition or by the experiment control condition were found. Similarly, no significant main effects of the material or experimental control conditions were found on the written tests.

In addition to the previous studies, a number of other studies have explored the comparative effects of TUIs against GUIs, yielding results which suggest TUIs offer a number of auxiliary benefits, such as increased engagement, fun, and collaboration. For example, Horn et al. carried out a study [18] which compared GUI and TUI implementations of their Tern programming language [17] in a public museum setting. They found that the TUI implementation was more inviting, more supportive of active collaboration, and more child-focused, though they found no differences in engagement. In another study, Rogers et al. [36] explored how transforms (i.e. the relationship between physical or digital actions and their physical or digital effect) can affect the learning process in young children. They found that a physical action leading to a digital effect, or vice versa, led to increased reflection and interest. Finally, a comparative study by Xie, Antle and Motamedi [43] explored the relationship between TUIs and the effect on enjoyment and engagement. In the study,
school aged children put together puzzles in one of three interfaces: GUI, physical (i.e. a standard puzzle), and TUI. They found that the children had more difficulty in completing the puzzles in the GUI condition and that the physical and TUI conditions were more engaging.

2.1.4 Discussion

Given the strong foundational backing to the notion that tangible interfaces ought to offer measurable benefit over their more traditional counterparts, it is rather surprising that a number of studies which have empirically tested this notion have found conflicting results. Most importantly, this calls into question whether there is any real world benefit to tangible interfaces on learning tasks. Given the variety of the learning tasks carried out among the studies, perhaps the effect of tangible interfaces is task dependent. If so, the necessity of carrying out more empirical investigations to discover which types of learning tasks benefit from a tangible interaction is even more important. The study of this thesis will be addressing this deficiency through its exploration of the effect of interface closeness on abstract problem solving and learning.

The presence of conflicting results highlights the importance of viewing the previous research in context. Broadly, the research in this area is investigating the role the interface and interaction plays in learning. This general area, however, is too broad. Learning has strong individual differences, especially across ages, so it is necessary to keep the demographics of the studies in mind. Of the studies which have found a significant benefit of TUIs over GUIs, most were carried out in children. As the participants in the comparative studies get older, a trend of the effect of the interface weakening appears. On the surface, this would suggest that further investigation should focus on the
role TUIs play in children, forgoing further investigation into TUIs in adults. It could be argued, however, that the beneficial effects of TUIs have yet to materialize in adults due to the lack of attention they have received in the community. Further, as previously mentioned, the effect of TUIs may be even more task-dependent in adults, thus emphasizing the need for more further investigation.

Similarly, because these studies were not carried out to investigate the effects of closeness, their comparisons focus less on the interfaces and task type, and more on the learning outcomes and practicality in the classroom. As a result, many of the studies utilize tasks which do not transfer well between interfaces, which lead to a biasing effect of one interface over another. The study of this thesis utilizes a task which transfers well between interfaces, thus eliminating such biases from the result. Along the same lines, previous research lacks the comparison of interfaces at each level of closeness (i.e. mouse, touch and tangible). This lack of comparison may suggest that researchers believe that there is no difference in closeness, but that only physicality has an effect, even though previous research [15] does not validate this belief. The proposed study will carry out the comparison study on all three levels of closeness, thus providing a more solid foundation from which to draw conclusions.

### 2.2 Intersection of Interface and Problem Solving

Problem solving, at its core, is the act or process of applying rules, techniques or strategies in order to find a solution to a given problem. The application of such techniques or strategies, however, is context dependent. Similarly, which strategies are chosen and how they are applied to a given
situation vary greatly across people, and even within a single person. Given the fluid nature of problem solving, it is important to understand how problem solving strategies can be influenced. One area in problem solving research that is of particular interest is the effect of the interface and interaction style on problem solving.

Svedsen’s *The influence of interface style on problem solving* [39] appears to be the first empirical investigation of its kind into the influence of the interface on problem solving. The study is built upon the work of Hayes and Broadbent [14], which states that there are two modes of learning, S-mode and U-mode. In S-mode, “learning takes place by means of abstract working memory and is selective and reportable.” In contrast, U-Mode “learning occurs outside abstract working memory and is unselective and unavailable for verbal report.” Svedsen, however, simplifies S-mode to be synonymous with higher order discovery learning, and relegates U-mode to mere trial and error. Applying these ideas to HCI, it was shown through two experiments that command-line interaction will induce S-mode learning, while direct manipulation will induce U-mode. Reviewing the first experiment, participants were required to solve the Towers of Hanoi puzzle in either a command-line or direct manipulation interaction. The Towers of Hanoi puzzle consists of three rods and 5 disks of graduating size which slide on the rods. The goal is to move the disks from one rod to another while following a number of rules. These rules are: only one disk can be moved at a time, only the top disk on each rod can be moved, and a larger disk cannot be placed on a smaller disk. Participants in the command-line condition completed the task, which required two-error free completions, in significantly less trials than the direct manipulation condition. Similarly, participants in the command-line condition also made significantly fewer errors than those in the direct
manipulation interface. Perhaps unsurprisingly, across both experiments, participants overwhelmingly preferred the direct manipulation interaction, due to increased ease of use.

Though the results of Svedsen’s experiments are compelling, the underlying reasons for the results are unclear. Fortunately, O’Hara and Payne’s *The Effects of Operator Implementation Cost on Planfulness of Problem Solving and Learning* [34], offers a strong explanation. Unlike Svedsen, O’Hara and Payne focus on the underlying mechanisms that may lead to the significant differences between command-line and direct manipulation in problem solving tasks. Their analysis centers on the cost of performing an action, and how that affects planfulness—a term they introduce to describe the level of planning during problem solving. Along these same lines, they introduce the notion of implementation cost, which “is the cost associated with bringing about the effects of a particular operator in the world” and includes factors such as time, as well as physical and mental effort. Within the framework of implementation cost, O’Hare and Payne explain Svedsen’s results as follows:

…the differential effects of the two interfaces might be explained by their different operator implementation costs. In the direct manipulation condition there is a low implementation cost associated with each operator; the command-based interface incurs a considerably higher implementation cost. Thus, by the cost-benefit argument, users of the command-based interface will be prompted to think harder, making correspondingly fewer errors and learning more efficiently.
This analysis of Svedsen’s work led to a series of experiments to further explore the effect of implementation cost on problem solving. Following in the vein of Svedsen’s comparison between command-line (high implementation cost) and direct manipulation (low implementation cost), O’Hare and Payne carry out a similar comparison which utilizes an 8-puzzle. The 8-puzzle consists of eight numbered tiles arranged in a 3x3 matrix, with one cell left empty. The goal of the puzzle is to rearrange the tiles, utilizing the empty cell, until the goal configuration is reached. The results corroborated and validated their cost/benefit analysis, with the high cost group making significantly fewer moves to reach the goal configuration.

Further, Noyes and Garland [33] conducted a number of experiments which investigated the effect of presentation on solving the Tower of Hanoi puzzle. Using a counterbalanced, within-subjects design; participants were to solve the puzzle in three interaction conditions: mental, computer and physical. In the mental condition, participants had no physical aids at their disposal; they had to solve the task entirely in their minds. In the computer condition, participants were afforded a direct manipulation interface, whereby they could ‘drag and drop’ the disks of the puzzles and see the effect of each action. Finally, in the physical condition, participants were given a paper model of the task which they could manipulate freely. Between the three conditions, participants in the mental condition were found to be significantly more efficient in completing the puzzle, though they also had significantly longer completion times, as well as a greater probability of failure. Participants in the computer condition had a significantly higher success rate, and shorter completion times, yet they generated more moves in order to solve the puzzle. Finally, participants in the physical condition had the highest rate of unsuccessful completions and, similar to their
computer condition counterparts, generated significantly more moves to solve the puzzle and had shorter average time per move.

2.2.1 Discussion

A common theme across these studies is that in order for an interface or interaction to increase problem solving ability, it is imperative for the interface foster reflection and planning. Adopting the verbiage of O’Hare and Payne, an interface which has a higher operator implementation cost will foster such problem solving characteristics. As a result, direct manipulation and physical interfaces, which have lower operator implementation cost, have yielded less efficient, though quick, problem solving. This would imply that such interfaces promote U-mode learning, or trial and error, as Svedsen simplified it. In contrast, command-line interfaces and mental interactions, which have higher implementation costs, have yielded the most efficient, though slow and error-prone, problem solving. These results seem to suggest that TUIs would not be beneficial in promoting efficient problem solving. Indeed, in Do tangible interfaces enhance learning? [24] Marshall argues that “it is possible that if tangible interfaces support easy manipulation of concrete objects, that they could in turn lead to decreased reflection, planning and learning.”

As was the case with early research in TUI for learning, however, such claims require empirical investigation. Thus far, command-line and direct manipulation interfaces have garnered significant research, while closer interfaces, such as touch and TUIs, have been virtually ignored by the community, despite previous research [15] and [9, 33], which suggest closer interactions do foster more efficient and faster performance over traditional interfaces. Given the previous discussion
of the possible learning benefits of closer interfaces, specifically TUIs (see Section 2.1.2), and the open question as to whether or not such benefits are task-dependent, it is important to investigate the interaction between closer interfaces and problem solving. This thesis will carry out a comparative study focused on abstract problem solving and learning on interfaces at three levels of closeness (i.e. mouse, touch and tangible), thus providing the necessary empirical results from which to decide if closer interfaces foster efficient, reflective problem solving.
CHAPTER 3

TASK AND SYSTEM DESCRIPTION

3.1 Task

Abstract problem solving and learning are facets of many scientific disciplines. In order to investigate the effects of interface closeness on these aspects, a task was required that fulfilled two main constraints. First, the task would need to center on abstract problem solving and learning, while also being engaging. Second, the task should lend itself to being implemented as both a GUI and a TUI, with little lost between each implementation. The board game Mastermind\textsuperscript{TM}, manufactured by Hasbro, a two-player code-breaking game that relies heavily on deductive reasoning as well as hypothesis generation and modification, worked as an ideal starting point. In Mastermind\textsuperscript{TM} the codemaker creates a code that needs to be discovered by the codebreaker. A code consists of four colored pegs placed in four positions placed in a row. Once the code has been set by the codemaker, the codebreaker makes an initial guess as to what the code is. The codebreaker’s guess is then analyzed by the codemaker in two ways: how many of the pegs in the guess are correct in both color and position and how many of the pegs are correct in just color. The codemaker then provides the codebreaker with feedback that consists of the results of their analysis of the current guess. Given the
feedback, the codebreaker must then revise their guess and the process will continue until the code has been broken.

3.1.1 Task Description

The task of our experiment retains the underlying feedback structure of Mastermind™ that fosters deductive reasoning and hypothesis generation, but it is important to note that they are distinctly different tasks. In the experimental task, each trial has a contiguous arrangement of blocks that needs to be discovered. In order to discover this arrangement, participants add, remove and rearrange colored blocks on a 4x4 grid. Where the arrangement takes place on the grid is not important, what is important is how the pieces are arranged relative to one another. Each trial’s arrangement has 3 core features; dimension, number of blocks present, and number of colors present.

Dimension refers to how many dimensions the trial’s correct arrangement of blocks exists in and can be either one or two. If the arrangement is wholly horizontal (i.e. it takes place all in one row) or wholly vertical (i.e. it takes place all in one column), it would be one-dimensional. If the arrangement were both horizontal and vertical (i.e. it takes place in two or more rows and columns simultaneously), it would be two-dimensional. The number of blocks present refers to how many blocks the trial’s arrangement consists of. This ranges from 2 to 4 blocks and 3 to 5 blocks, for one and two dimensional configurations, respectively. The number of colors present refers to how many colors the trials correct arrangement consists of. This ranges from 1 to 4 colors and 2 to 5 colors, for one and two dimensional configurations, respectively. Overall, there are 5 colors available: white, blue, red, yellow and green. A trial’s arrangement can be thought of as a number of pairs put
together in a specific way. So, for each arrangement, there are a set of pairs that it can be
decomposed into. There are two types of pairs: horizontal and vertical. An example of a pair would
be a blue block to the left of a red block. It should be noted that the reverse of this pair, a red block
to the left of a blue block, is a different pair. It should be noted that two or more pairs can share a
common piece (see Figure 3.1a for an example).

For each trial, the correct arrangement’s dimension, number of blocks and number of colors
is constantly available to the participant. Initially given only the core features of the arrangement, the
participant will begin each trial by making an initial submission. With each submission, the system
will analyze it and give two types of feedback about the submission to the participant. The first is
how many colors the participant’s submission has in common with the trial’s arrangement. The
count is context-free—where the color block is and how many blocks of the color are present on the
grid have no bearing on this aspect of the feedback given. The second is how many pairs the
participant’s submission has in common with the correct arrangement. Just as the trial’s arrangement
is decomposed into pairs, so is the participant’s submission. The number of pairs the participants
submissions has in common with the trial’s arrangement will be reported back to the user. Multiple
instances of the same pair are ignored by the system, unless the trial’s arrangement has multiple
instances of that specific pair. Finally, just like the feedback regarding how many colors were correct,
where the pair is positioned on the grid is irrelevant to the feedback and to the correctness of a
submission.
Figure 3.1: (a) Example trial arrangement from the experimental task
(b - g) Example solution path taken by a participant

A complete example of a trial found in the experimental task, including the trial’s arrangement and an actual solution path by a participant in the study is shown in Figure 3.1. In Figure 3.1a, the arrangement for a given trial in the experimental task is shown, and must be discovered and submitted by the participant. The core information for this trial is the following: 2 Dimensions, 3 Blocks, 2 Colors. This arrangement would be decomposed into two pairs: a red block on top of a blue block, and a blue block to the left of a red block. The initial submission of the participant is shown in Figure 3.1b. Upon submission, the system would report that there were 1 correct colors and 0 correct pairs, and would repeat that feedback for the following two submissions.
shown in Figure 3.1c and d, respectively. The submission shown in Figure 3.1e would yield feedback of 2 correct colors and 1 correct pairs, as would the submission found in Figure 3.1f. With the submission of the configuration shown in Figure 3.1g, the participant has discovered the trial’s arrangement and will proceed to the next trial.

The decision to include the number of colors correct in a given submission instead of forcing the participant to deduce the colors through the pair feedback was born from the tasks’ evolution from Mastermind, where the color feedback is given as a separate, weaker piece of information. Once the participant discovers the correct arrangement for the trial, the next trial will immediately begin. There were 32 total experimental trials available to be completed.

3.1.2 Arrangement Creation

The arrangement for each trial was generated within a number of constraints. The first constraint was that as the participant worked through the trials, the complexity should gradually increase. Complexity refers to how difficult a given trial’s arrangement is to discover. The gradually increase was designed to minimize the risk of a participant getting stuck on a difficult trial early on, thus limiting the amount of data gathered. In order to accomplish this, the complexity of the arrangements needed to be quantified. As a simple solution, complexity was defined as the product of the number of three core features of every solution (i.e. Dimension x Number of blocks x Number of colors). Another constraint was the board size. Given the grid is of size 4x4, this automatically capped the one-dimensional arrangements at a maximum of 4 blocks. Thinking of each trial’s arrangement as a set of pairs put together in a specific way, this sets the minimum
number of blocks possible for an arrangement to a given trial at 2. With the block range for one-dimensional solutions being 2 to 4, the only other parameter left to change was the number of colors. For all arrangements with more than two blocks, the one color arrangement was discarded. The decision to discard the one color option for arrangements of more than two blocks was made because it offered little opportunity for problem solving and learning, which would make such trials essentially worthless. Finally, two-dimensional arrangements were capped at a block size of five as a result of pilot testing, which showed that adding any more levels of complexity, and thus more trials, would yield diminishing returns, as few pilot participants made it to the final trials within an hour time limit. Each of the thirty-two trial’s arrangements can be found in Appendix A.

3.1.3 Task Benefits

The experimental task was designed to test problem solving within a broader context of leveraging aspects of discovery learning. There are two phases of the experimental task: discovery and synthesis, which leverage different aspects of discovery learning. During the discovery phase, the participant searches to establish both the colors present in the task, and what pairs compose the trial’s arrangement. The feedback given to the participant as a result of their submissions will build a foundation of knowledge about the trial’s arrangement that is constantly updated. The synthesis phase is characterized the process of combining and testing the information acquired in the discovery phase. It is during this phase where the hypothesis generation and modification are most present. As the participant works, the number of correct pairs will rise and fall, forcing constant revision and expansion upon a current hypothesis, or to devise a completely new one. It is important to note that
these two phases are not discrete (i.e. aspects of each phase overlap somewhat) nor linear (i.e. the participant will move back and forth between the phases).

3.2 Software

The experimental task, as outlined in Section 3.1.1, was implemented in the Java programming language using Oracle’s open-source Netbeans IDE. Java was selected because the library for TopCodes [16], which are barcode like symbols detectable by a standard webcam, is written in Java, and the TopCodes were chosen to use in the implementation of the tangible interface condition. The software presented the user with a 4x4 grid (in the middle), 5 colored blocks (on the right side), trial information (above the grid) and a submission button (below the grid) (see Figure 3.2). In the mouse and touch condition, the colored blocks on the side can be interacted with by clicking/touching and dragging them onto and around the 4x4 grid. In order to make each submission, the user will click/touch the submit button.
3.3 Physical Setup

The study was performed at Bowling Green State University in the Computer-Human Interaction Lab (CHIL), located at 227 Hayes Hall. Participants were seated in front of the workstation and there were three different interaction types: mouse, touch and tangible, which had different equipment. All participants viewed the software on a GVISON P15BX 15-inch XGA 1024 x 768 resolution touchscreen monitor. The monitor’s touchscreen panel is a 5-wire resistive, capacitive monitor with a typical response time of 8 milliseconds. Participants in the mouse interaction condition utilized a standard Dell mouse to interact with the software’s GUI. Participants in the touchscreen condition utilized the GVISON monitor’s touchscreen capability to interact with the software’s GUI.
3.3.1 Tangible Condition

In front of the user is a base, created from Lego™ blocks. The base is the physical representation of the 4x4 grid present in the software. In order to add and remove blocks from the grid, participants will physically add and remove the colored blocks available to them. Like the base, these blocks are also composed of Lego™ blocks which allow them to be stacked as in the software. The 4x4 grid in the software will constantly update to reflect the blocks that are present on the base. In order for the software to know what blocks are present on the base, each of the colored blocks has been augmented with TopCodes, which were detected using a Logitech webcam which faces the back of the base. In order to make a submission, the participant has a submission block that they place in the ‘Submit’ position of the base and then remove it. The act of adding and removing the submission block is meant to replicate the submission button in the software (which is not present in the software for the tangible condition), without changing the mode of interaction for the participant. There are a total of 6 TopCodes present in the system, one for each color and one for the submission block. The tangible condition, with an example set of blocks placed on the base is shown in Figure 3.3.
Figure 3.3: Setup of the tangible condition
CHAPTER 4

METHODS

4.1 Introduction and Hypotheses

This chapter will describe the study conducted to test the effects of interface closeness on abstract problem solving and learning. The primary goal of this study was to determine whether or not a closer interface (e.g. touch or tangible) lead to cognitive benefits over an interface that is less close (e.g. mouse). A secondary, more specific goal was to determine if the physicality of an interface, present in tangible UIs, will lead to cognitive benefits. There were several hypotheses: (H1) Participants in the tangible condition will have the lowest average number of submissions of the three interfaces, (H2) Participants in the tangible condition will have the longest average submission time, (H3) There will be no significant differences by condition in overall number of completed trials, (H4) There will be no significant differences between participants in the mouse and touch conditions across all measures. Each hypothesis has a corresponding null hypothesis, referred to as HxN, where x is 1, 2, 3 or 4. For H1 and H2, the null hypotheses state that there are will be no significant differences found. In contrast, the null hypotheses for H3 and H4 state that there will be significant differences found.
4.2 Experimental Design

The study was comprised of three conditions utilizing a between-subjects design. The independent variables were the closeness of the UI (with mouse, touch and tangible composing the three levels), which was between subjects and the complexity of the trials, which was within subjects. Dependent variables included: number of trials completed overall, time taken per trial, time taken per submission within each trial, number of submissions made per trial, and number of blocks used per trial. All timing-related measures were recorded in millisecond time.

4.3 Participants

Forty-three college student participants (37 men, 6 women) with a mean age of 22.3 years old volunteered for this study. Participants were recruited from upper-level Computer Science classes (Computer Graphics, Usability Engineering, and Information Management Technologies) at Bowling Green State University. In return for their participation, their instructors offered them extra credit in their respective classes.

4.4 Experimenters

The author and a fellow graduate student were the only two experimenters. An experimental procedure sheet, which included a script that was read to each subject, was utilized to limit bias and to ensure that each experimental session was run consistently and correctly. Both experimenters completed the required HSRB training before running any participants.
4.5 Procedure

After each participant arrived at the lab they were given the consent form (see Appendix B) and asked to read it before signing it. Once consent was received, the participant filled out the demographics form (see Appendix C). While the participant filled out the form, the experimenter assigned the participant to one of the three UI conditions (mouse, touch or tangible). The assignments were done randomly to guard against bias. Once the demographics form was filled out, the experimental portion of the study began only if the participant had reported that they were not color blind.

To ensure that each participant understood the task, how to interact with the system, and the feedback they would be receiving throughout, a training session preceded the experimental trials. First, the experimenter gave a thorough description of each aspect of the experiment. This included an explanation of what the goal each trial was, how to interact with the software through their given UI condition, as well as how to understand the feedback they would receive from the system with each submission. Once the explanation was complete, each participant began the training trials. The only difference between the training and experimental trials was that the experimenter was allowed to answer any questions that the participant may have had and offer help if needed. There were three training trials, chosen to represent each type of correct arrangement that would be found in the experimental trials (i.e. a one-dimensional, one color arrangement; a one-dimensional, multiple color arrangement; and a two-dimensional, multiple color arrangement). Each participant was allowed to take all the time needed to complete the training trials.
Once the training trials were completed, the experimental trials directly followed. There were a total of 32 experimental trials, and participants had one-hour to complete as many of these trials as they could. If a participant had not completed all trials within the one-hour limit, the software stopped and the participant’s time working on the experimental trials was finished. During the experimental trials, the experimenter was allowed to answer only questions relating to how to interact with the system.

After the experimental trials, participants were given the NASA Task Load Index (TLX) [31] (see Appendix D) questionnaire to report the workload that the experiment placed on them. Once completed, the participants had completed all aspects of the experiment.

4.6 Data Collection

The data collected for the study are drawn from three sources: the demographics form, which captured basic information about each participant; training and experimental trial data, which comprised the dependent variables of the experiment and was collected by the system; and the NASA TLX questionnaire, which captured workload as reported by the participant.

4.6.1 Demographics

Before the experiment began all participants were asked to fill out a demographics form. The information collected from the form was: age, sex, experience with tangible interfaces (such as tablets, touch screen devices, Nintendo Wii remote, etc.) and handedness. The demographics form can be found in Appendix C.
4.6.2 Training and Experimental Trials

All data from the practice and experimental trials were collected from the Java program driving the experiment. Screenshots of each submission were also recorded. The screenshots play an important role during analysis as they allow for reconstruction of whole trials, which gives a look inside the participants mind as they were attempting to discover the solution—something the performance data alone does not afford.

4.6.3 TLX Questionnaire

After the participants completed their experimental trials, they were asked to fill out a NASA Task Load Index. This questionnaire measures the workload placed on the participant during the experiment on six different categories: mental demand, physical demand, temporal demand, performance, effort, and frustration. Each of these measures range on a scale from 1 to 20. The NASA TLX questionnaire can be found in Appendix D.
CHAPTER 5

RESULTS

This chapter reports the results of the study previously described in Chapters 4. A total of 43 subjects were run of which 42 produced viable data. One subject was removed from the analysis as a result of quitting the experiment. The results will be broken into two sub-sections: analysis of performance, and analysis of workload. The analysis of performance covers all objective performance related measures. Broadly, these have been divided into two categories: speed, and efficiency. Speed centers on measures related to the speed at which participants were able to complete the problem solving task, while efficiency focuses on measures related to how efficient participants accomplished the problem solving task. The analysis of workload covers all measures related to the self-report NASA TLX questionnaire (see Section 4.6.3) data. The analyses carried out on both types of measures included: multiple analyses of variance (MANOVAs), post-hoc pairwise comparison testing utilizing Tukey’s-HSD, and pairwise bivariate correlational testing.

5.1 Analysis of Performance

For the analysis of performance, all but one comparison was based on the data from the first 24 trials, as completion of 24 out of a possible 32 trials was established as the baseline for comparison. This resulted in 3 subjects being removed from the analysis, leaving 39 participants overall with 13
in each condition. The one comparison utilizing all 42 participants was that of the dependent variable Number of Trials Completed Overall.

5.1.1 Speed

There were two performance measures related to speed: Average Time Taken per Trial (measured in seconds) and the aforementioned Number of Trials Completed. There was a significant main effect of Average Time Taken per Trial, \( F(2,36) = 5.07, p < .05 \). For Average Time Taken per Trial, post-hoc comparisons revealed significant differences between the Tangible condition (\( M = 107.02, SD = 21.91 \)) and both the Mouse condition (\( M = 81.35, SD = 24.97 \)) and Touch condition (\( M = 77.95, SD = 28.99 \)), see Figure 5.1.

![Figure 5.1: Average Time Taken per Trial](image)

For Number of Trials Completed, no significant differences were found between the Tangible (\( M = 27.36, SD = 3.30 \)), Mouse (\( M = 30.00, SD = 3.76 \)) and Touch conditions (\( M = 29.86, SD = 3.94 \)).
5.1.2 Efficiency

There were 3 performance measures related to efficiency: Average Number of Blocks Used per Trial, Average Number of Submissions per Trial, and Average Time Taken per Submission (measured in seconds). There were significant main effects of Average Number of Blocks Used per Trial, $F(2,36) = 8.22, p < .01$; Average Number of Submissions per Trial, $F(2,36) = 5.45, p < .01$; and Average Time Taken per Submission, $F(2,36) = 19.38, p < .001$. For Average Number of Blocks Used per Trial, post-hoc comparisons revealed significant differences between the Tangible condition ($M = 34.96$, $SD = 6.73$) and the Mouse condition ($M = 62.13$, $SD = 26.14$), as well as between the Touch ($M = 43.91$, $SD = 13.45$) and Mouse conditions, see Figure 5.2.

![Figure 5.2: Average Number of Blocks per Trial](image)

For Average Number of Submissions per Trial, post-hoc comparisons revealed significant differences between the Tangible condition ($M = 13.12$, $SD = 1.52$) and the Mouse condition ($M = 18.76$, $SD =
6.99), as well as between the Touch Condition ($M = 13.95, SD = 3.89$) and the Mouse condition, see Figure 5.3.

![Figure 5.3: Average Submissions per Trial](image)

Finally, for the Average Time Taken per Submission, post-hoc comparisons revealed significant differences between the Tangible condition ($M = 8.15, SD = 1.27$) and both the Mouse condition ($M = 4.53, SD = 1.06$) and Touch condition ($M = 5.78, SD = 2.02$), see Figure 5.4.

![Figure 5.4: Average Time Taken Per Submission](image)
5.2 Analysis of Workload

For the analysis of workload, all 42 participants were present in the analysis. There was a significant main effect of the Frustration measure $F(2,39) = 3.78$, $p < .05$. A post-hoc comparison revealed a significant difference between the Tangible condition ($M = 7.93$, $SD = 1.35$) and the Touch condition ($M = 12.79$, $SD = 1.35$). Further, post-hoc comparisons revealed a significant difference between the Tangible condition ($M = 10.36$, $SD = 3.71$) and the Mouse condition ($M = 6.07$, $SD = 4.16$) for the measure of Temporal Demand, see Figure 5.5. Additionally, the Effort measure was significantly positively correlated with both the Frustration, $r(42) = 0.56$, $p < .001$, and Mental Demand, $r(42) = 0.38$, $p < .05$, measures. A significant positive correlation between the Physical Demand and Temporal Demand measures, $r(42) = 0.4$, $p < .05$, was also revealed.

![Figure 5.5 Summary of results from the NASA TLX](image-url)
CHAPTER 6

DISCUSSION AND FUTURE WORK

In this chapter, the overall findings are summarized as they relate to the hypotheses of the study, followed by a discussion of the results found in Chapter 5. Further, a discussion of the limitations and avenues for future research are explored. Finally, a section considers the overall implications, contributions and conclusions to be drawn from this work.

6.1 Overall Findings

Restating the hypotheses: (H1) Participants in the tangible condition will have the lowest average number of submissions of the three interfaces, (H2) Participants in the tangible condition will have the longest average submission time, (H3) There will be no significant differences by condition in overall number of completed trials, (H4) There will be no significant differences between participants in the mouse and touch conditions across all measures. Reviewing each hypothesis, H1, was partially supported by the findings. Participants in the tangible condition averaged significantly fewer submissions than those in the mouse condition, and had the lowest number overall, but not significantly less than those in the touch condition. H2 was supported by the findings. Participants in the tangible condition averaged significantly longer time per submission than participants in both the mouse and touch conditions. H3 was also supported, as there were no significant differences in
the overall number of trials completed by condition. Finally, H4 was not supported by the findings. On two performance measures, average number of blocks used per submission, and average submissions per trial, participants in the touch condition averaged significantly fewer than those in the mouse condition.

6.2 Discussion

6.2.1 Speed and Efficiency

Previous research into the intersection of interfaces and problem solving (see Section 2.2) have shown that increased implementation cost led to increased reflection and efficiency in problem solving, though speed of problem solving decreased. The significant differences between conditions on the experimental task, however, are contradictory to the previous research. For example, participants in the tangible condition had significantly longer average time taken per trial than both the mouse and touch conditions. In the context of previous research, this result would only make sense if the tangible condition had a higher implementation cost. However, in terms of interface closeness, the tangible condition should have the lowest implementation cost, as it is the closest of the three interface conditions. Furthermore, in terms of problem solving efficiency, participants in both the tangible and touch conditions were significantly more efficient than those in the mouse condition. Once again, this appears contradictory to the result that increased implementation cost should lead to more reflective and efficient problem solving.
Though the underlying causes are unclear, there are a number of possible avenues which may explain the results of the tangible condition. One possible explanation could be that the effect of the closeness of the interface was canceled out by the novelty of the tangible interface. Though all participants were given an in-depth explanation and completed a set of training trials in their interface condition, for many users, the experimental task may have been their first experience with a tangible interface. Therefore, as most users have spent many years with mouse and touch interfaces, but only a few minutes with the tangible interface, some performance deficits should be expected. Another explanation could be found in previous research [36], which showed that physical actions leading to digital effects led to increased reflection. Finally, it was observed that some users utilized the physical nature of the blocks to organize their thoughts by working with the blocks off the grid and without submitting (though it is possible that this is a symptom of increased reflection time, and not the cause).

The results of the touch condition are less clear. The touch condition is both as fast as the mouse condition and as efficient as the tangible condition. As was the case with the tangible condition, the touch condition is closer than the traditional mouse interface. According to the predictions of the work into implementation cost, the touch condition should be fast, though inefficient. Unlike the tangible condition, however, the idea that the effect of closeness was offset by the novelty of the touch interface does not apply, as touchscreen have seen widespread use with the explosive growth of the smartphone and tablet computer markets. One possible explanation for the increased reflection and efficiency may be a product of the touch condition’s increased concreteness over that of the mouse condition. In contrast to the tangible condition, which has physical
manipulatives, the touch condition has virtual manipulatives. As previously discussed (see Section 2.1.2), virtual manipulatives have been shown to speed up the transfer of from manipulative to abstract mental representation. It is possible, therefore, that the touch condition’s more concrete nature allowed participants to move more fluidly between what they were manipulating on the grid, and the abstract problem solving they were doing simultaneously.

Finally, the results of the mouse condition are less surprising. Like the results of the tangible and touch conditions, those of the mouse condition appear to contradict the predicted results when viewed through the lens of implementation cost. However, it is important to remember the overall context in which the three interfaces were tested. While the mouse condition is least close of the three interfaces used in this study and thus the highest implementation cost of the three, to the user, the mouse is the default manner by which they have interacted with computers for decades. As a result, the relative increase in implementation cost in comparison to the other two interfaces may be cancelled out due to the user’s familiarity. If this were indeed the case, previous research [39] suggests that the performance data should reflect a trial and error approach. A trial and error approach would be characterized by shorter submission times, higher number of blocks used per submission on average, and higher number of submission per trial on average—all of which match up to the performance results of the mouse condition. Therefore, the relative increase in implementation cost in comparison to the other two interfaces was offset by the users’ familiarity.
6.2.2 Workload

By definition, interfaces which are closer are more natural, and should therefore be less frustrating. Tangible interfaces are inherently closer than touchscreen interactions. As a result, it is not surprising that participants in the tangible condition rated their frustration level significantly lower than those in the touch condition. This line of argument is supported by the difference in rating between the tangible and mouse conditions which, though not significant, was trending in that direction. It should be noted that the frustration ratings of participants in the mouse and touch conditions could have been influenced by an interface bug in which the system believed a block was present where there wasn’t one. This system bug was present in the mouse and touch conditions only. These “ghost blocks,” appeared relatively infrequently and were easily identified and removed, but some users reported their presence as frustrating.

The significant difference in the temporal demand ratings between the tangible and mouse conditions, with participants in the tangible condition averaging significantly higher ratings, is surprising. The TLX questionnaire (see Appendix D) frames the temporal rating with the following question, “How hurried or rushed was the pace of the task?” It was not predicted that there would be any differences in ratings, but in hindsight, the result appears to go against the performance results. By all objective measures, participants in the mouse condition moved through the experimental task much more quickly than their counterparts in the tangible condition—significantly so. One possible explanation for this result could be an interaction between multiple factors. To review the experimental design, participants were given one hour to complete as many of the 32 experimental
trials as they could. This time limit, however, was not made known to them, and in fact users were explicitly told to work at whatever pace felt comfortable for them. At the end of the hour, if the participant had not completed all 32 trials, the system would notify the user that their time was up, and then the experimenter would have them fill out the TLX questionnaire. Though there was not a significant difference found by condition in the overall number of trials completed, more participants in the mouse condition than in the tangible condition completed all 32 trials within the hour. Thus, with the resulting system message notifying them that their time was up still fresh in their mind, which perhaps lead to a belief that they had not performed the task fast enough, may have primed them to rate that the task was more hurried and rushed, due to their perceived slow performance. It should be noted, however, that the participants had no idea how many trials were present in the experimental task, how long the time limit was, and were not given any indication of how well or poor they did, therefore any perceived notion of a “slow” or “poor” performance would have been of their own creation.

When designing the task, one of the core constraints was that the task needed to be engaging. This was a necessity, as the task was meant to investigate discovery learning (see Section 1.2), and the core tenet of discovery learning is that the learner will construct their own knowledge through self-guided and self-motivated exploration. If the task was not engaging, it would be ill-suited for discovery learning—defeating the purpose. Reviewing the correlational data, the significant positive correlations between effort and mental demand, reveal a pattern which suggests engagement. That is, though participants found the experimental task to be mentally demanding, they rose to the challenge because they were engaged in the task at hand, despite its difficulty.
Further, the significant positive correlation between effort and frustration suggests engagement, as it seems less likely that someone would be frustrated about a given situation unless they had some sort of vested interest in its outcome. Thus, in the case of the study, participants were frustrated because they actually cared about their performance and about solving the puzzles that comprise the task. Finally, the significant positive correlation between physical demand and temporal demand makes little sense on its own. However, in the context of the previous result centered on temporal demand ratings, it may offer some insight as to why participants in the tangible condition had significantly higher ratings than those in the mouse condition. Though there were no significant differences in physical demand ratings by condition, the physical nature of the tangible condition could have made some impact on the temporal demand ratings.

6.2.3 Complexity

As previously discussed in Section 3.1.2, in order for trials to increase in complexity as the task progressed, complexity needed to be operationalized. For this study, complexity was defined as the product of three components of each trial’s correct configuration: Number of Dimensions, Number of Blocks, and Number of Colors. By definition, therefore, it is assumed that each of the three components contribute equally to the complexity of a given block configuration. Upon a qualitative analysis however, this assumption appears unlikely. Ranking trials based on a composite score of average time taken per trial, average number of submissions per trial and average number of blocks per trial, the six most difficult trials, across conditions, were (from most complex to least): 21, 28, 12, 29, 30, and 25 (see Appendix A). After decomposing each of these trials into their respective
components, the Number of Dimensions and Number of Blocks were strong contributors to the
complexity of a given configuration, as 83% were 2-Dimensional, 67% were composed of 5 blocks.
Another factor, however, was the strongest single contributor to complexity—repetitious colors.
When the Number of Colors is less than the number of blocks, one or more colors will have to
repeat. All six of the most complex shapes contained repetitious colors; 83% of those only had one
color repeating, the remaining had two colors repeating (the maximum number of repeating colors).
Therefore, any future work incorporating the experimental task introduced in this thesis should craft
a new definition of complexity primarily based on repetitious colors, number of dimensions, and
number of blocks. With that said, though the above recommendations would likely give a more
accurate operational definition, the interaction between complexity and performance on the
experimental problem solving task is not that simple. For example, as mentioned previously, by
design there are two trials at each level of complexity. Reviewing the list of the six most complex
configurations (i.e. trials 21, 28, 12, 29, 30, and 25) however, only two of them, 29 and 30, were
from the same complexity level. This would suggest that there are some more enigmatic elements of
complexity that are not easily operationalized and that further research is necessary.

6.3 Limitations and Future Work

Future work in this line of research must overcome three core limitations of the work presented in
this thesis. First and foremost, as the study is currently designed, it is virtually impossible to
understand and categorize the strategies used by participants during the experimental task. As is the
case in any problem solving task, there are any number of strategies available in order to discover the
correct configuration of each trial in the experimental task. While objective measures of performance are important in resolving how interface closeness affects participant performance on problem solving tasks, it gives no information on how interface closeness affects participant strategy. Further research should incorporate a think-aloud protocol into the design of the experiment. As their name implies, think-aloud protocols have the participant verbally share their thoughts, feelings, actions and explanations of those actions as they carry out an experimental task to an observer. The purpose of think-aloud protocols is to afford experimenters a look inside the mind of the participant as they perform a specific task, yielding a more complete picture of their performance. Incorporation of the think-aloud protocol in further research would lead to a better understanding of how interface closeness affects participant strategy.

Second, in the study’s current incarnation, there is no way to test the learning of the participants across the experimental task. Previous research [39] has said that different interfaces promote different types of learning. For example, interfaces with high implementation cost promote higher order discovery learning, while interfaces with low implementation cost promote mere trial and error. Though it was somewhat possible to categorize the three interface conditions into these categories (e.g. the mouse condition appears to have promoted trial and error), this categorization, however, does not necessarily mean that participants in one interface learned less than another. In order to test the amount learned across the experimental task, a sort of final task is required. Though there are a number of possible final tasks, one example is the following. After the participant completed some predetermined number of trials from the current experimental task (e.g. 24, based on the results of the current study), there would be a set of further experimental trials that needed to
be completed in a certain number of moves or less. With this final test, it would force participants to solve the problem in as few submissions as possible. For example, if the trial and error of the mouse condition led to reduced learning, and therefore a less complete understanding of how to solve the task, this should result in poorer performance on the final task. However, if participants in the mouse condition performed as well as participants in the other conditions, then the trial-and-error approach promoted by the mouse condition actually has no detrimental effect on learning, despite expectations to the contrary.

Finally, as previously discussed (see Section 6.2.3), the complexity measure utilized in this study did not accurately capture the complexity, in terms of difficulty to solve, of the trial configurations used. Based on the findings of a qualitative analysis, a number of recommendations were made so as to more accurately operationalize the concept of complexity as it relates to the experimental task proposed in this thesis. While the recommendations would serve as a better starting point, there were still a number of unanswered questions about the nature of complexity and its relation to performance on the experimental task which require further research.

6.4 Conclusion

This thesis has presented an empirical investigation conducted to determine the effects of interface closeness on abstract problem solving and learning. A secondary goal was to determine whether learning benefits for tangible interfaces may be task domain dependent. The study compared interfaces at three levels of closeness (mouse, touch and tangible) on a novel problem solving task which centered on deductive reasoning and discovery learning.
It was shown that touch and tangible interfaces can offer significant benefit over traditional mouse interfaces on an abstract problem solving and learning task. Individually, the touch interface yielded the best combination of speed and efficiency. The tangible interface was slightly more efficient than the touch interface, though also the slowest. However, it is suggested that the speed deficits of the tangible interface were a result of the novelty of the interaction. Finally, the mouse interface was as fast as the touch interface, though significantly less efficient than the other two. These results suggest that interfaces with concreteness and physicality foster reflection and planning, underlining their importance for interfaces designed specifically for problem solving tasks, as well as interfaces designed for discovery learning tasks in general. Furthermore, these results suggest that the learning effects of TUIs may be task domain dependent, highlighting the need for further research.
BIBLIOGRAPHY


[40] Triona, L. M., & Klahr, D. 2003. Point and click or grab and heft: Comparing the influence of physical and virtual instructional materials on elementary school students' ability to design experiments. Cognition and Instruction 21, 149-173.


APPENDIX A: TRIAL ARRANGEMENTS

Trial 1:

Trial 2:

Trial 3:

Trial 4:

Trial 5:

Trial 6:

Trial 7:

Trial 8:

Trial 9:

Trial 10:

Trial 11:

Trial 12:
Trial 13:

Trial 14:

Trial 15:

Trial 16:

Trial 17:

Trial 18:

Trial 19:

Trial 20:

Trial 21:

Trial 22:

Trial 23:
Trial 24:

Trial 25:

Trial 26:

Trial 27:

Trial 28:

Trial 29:

Trial 30:

Trial 31:

Trial 32:
APPENDIX B: CONSENT FORM

Consent to Participate in a Research Project
Bowling Green State University
Bowling Green, OH 43403

Hello, my name is Thomas Donaldine. I am a graduate student in Computer Science at Bowling Green State University. You are invited to participate in a research project about the effect of interface closeness on abstract problem solving and learning. Interface closeness is a term that describes how removed the user is from the data they are manipulating (e.g., mouse input is more removed than touch input). As part of our work on this project in the Department of Computer Science, we are conducting a study to help us understand how interface closeness can impact the learning of a user on an abstract problem solving task which isn’t typically thought to be impacted by interface closeness. In our research project, you will be asked to complete several trials of a problem solving task in which you will moving blocks in one of three interfaces (mouse, touch or tangible). You will also be asked to complete a questionnaire regarding your previous experience with tangible interfaces and to rate your experience after the experiment is completed.

You will be asked to participate in one session which will last no more than 1 hour and 30 minutes, most sessions will last no more than one hour.

The anticipated risks to you are not greater than those normally encountered in daily life.

The benefit to this project is that we can learn how to better utilize tangible user interfaces in problem domains not typically associated with tangible interfaces. There is, however, no direct benefit to you, the participant, from this project.

Information that you provide will remain confidential and your identity will not be revealed. Confidentiality of you as a respondent and your responses will be protected throughout the study and publication of study results. All of your responses will remain locked in a secure area within Hayes Hall, accessible only by Dr. Leverthol, the CS department chair, the CS department secretary, and the CS system administrator. Your identity will not be revealed in any published results.

Your participation in this study is completely voluntary, and you can refrain from answering any or all questions without penalty. If your instructor has agreed to provide extra credit for your participation in this study your instructor will be notified after the conclusion of the experiment and you will be awarded the amount previously agreed to by your instructor. If you decide to participate and change your mind later, you may withdraw your consent and stop your participation at any time without penalty. Your participation will have no impact on your compensation, your grades, class standing or relationship to Bowling Green State University. Again, you are free to withdraw consent and to discontinue participation in the project at any time.

If these conditions have been explained to you, you are willing to participate, you have had your questions answered, and you are age 18 or over, please sign below.

________________________________________________________________________
Signature

________________________________________________________________________
Date

If you have any questions or comments about this study, you can contact us

Thomas J. Donaldine or G. Michael Fosk, Visiting Assistant Professor
Computer Science Department
Bowling Green State University
Bowling Green, OH 43403
419-372-3337
donaldine@bgsu.edu

If you have questions about the conduct of this study or your rights as a research participant, you may contact the Chair of Bowling
Green State University’s Human Subjects Review Board at: (419) 372-7716

BGSU#024 - APPROVED FOR USE
IRB# # 312736
EFFECTIVE 04/16/2012
EXPIRES 03/31/2013
APPENDIX C: DEMOGRAPHICS

Participant #: _____  Group: _____

General

1. Age: _____

2. Sex:  Male      Female

3. Do you have experience with tangible interfaces (tablets, touch screen devices, Nintendo Wii remote, etc.)?
   Yes          No

4. If Yes:
   a. How much experience would you say that you have had with tangible interfaces?
      (1 – Not much, 5 – A great deal)
      1  2  3  4  5

5. Which hand is dominant?
   Left      Right      Neither (ambidextrous)

6. If Left or Neither:
   a. Do you typically use a left-handed mouse?
      Yes          No
APPENDIX D: NASA TLX

Figure 8.6

**NASA Task Load Index**

Hart and Staveland's NASA Task Load Index (TLX) method assesses work load on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.

<table>
<thead>
<tr>
<th>Name</th>
<th>Task</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mental Demand</td>
<td>How mentally demanding was the task?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Very Low</td>
<td>Very High</td>
</tr>
<tr>
<td>Physical Demand</td>
<td>How physically demanding was the task?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Very Low</td>
<td>Very High</td>
</tr>
<tr>
<td>Temporal Demand</td>
<td>How hurried or rushed was the pace of the task?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Very Low</td>
<td>Very High</td>
</tr>
<tr>
<td>Performance</td>
<td>How successful were you in accomplishing what you were asked to do?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Perfect</td>
<td>Failure</td>
</tr>
<tr>
<td>Effort</td>
<td>How hard did you have to work to accomplish your level of performance?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Very Low</td>
<td>Very High</td>
</tr>
<tr>
<td>Frustration</td>
<td>How insecure, discouraged, iritated, stressed, and annoyed were you?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Very Low</td>
<td>Very High</td>
</tr>
</tbody>
</table>