Supporting Collaborative Learning in Computer-Enhanced Environments

Shalva S. Landy

Graduate Center, City University of New York

How does access to this work benefit you? Let us know!
Follow this and additional works at: http://academicworks.cuny.edu/gc_etds

Recommended Citation
http://academicworks.cuny.edu/gc_etds/1015

This Dissertation is brought to you by CUNY Academic Works. It has been accepted for inclusion in All Graduate Works by Year: Dissertations, Theses, and Capstone Projects by an authorized administrator of CUNY Academic Works. For more information, please contact deposit@gc.cuny.edu.
SUPPORTING COLLABORATIVE LEARNING
IN COMPUTER-ENHANCED ENvironments

by

SHALVA S. LANDY

A dissertation submitted to the Graduate Faculty in Computer Science in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2015
This manuscript has been read and accepted by the Graduate Faculty in Computer Science in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Professor James Cox

Date

Chair of Examining Committee

Professor Robert M. Haralick

Date

Executive Officer

Professor Lori L. Scarlatos

Professor Zhigang Zhu

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK
Abstract

Supporting Collaborative Learning in Computer-Enhanced Environments

by

Shalva S. Landy

Advisor: James Cox

As computers have expanded into almost every aspect of our lives, the ever-present graphical user interface (GUI) has begun facing its limitations. Demanding its own share of attention, GUIs move some of the users’ focus away from the task, particularly when the task is 3D in nature or requires collaboration. Researchers are therefore exploring other means of human-computer interaction. Individually, some of these new techniques show promise, but it is the combination of multiple approaches into larger systems that will allow us to more fully replicate our natural behavior within a computing environment. As computers become more capable of understanding our varied natural behavior (speech, gesture, etc.), the less we need to adjust our behavior to conform to computers’ requirements. Such capabilities are particularly useful where children are involved, and make using computers in education all the more appealing.

Herein are described two approaches and implementations of educational computer systems that work not by user manipulation of virtual objects, but rather, by user manipulation of physical objects within their environment. These systems demonstrate how new technologies can promote collaborative learning among students, thereby enhancing both the students’ knowledge and their ability to work together to achieve even greater learning. With these systems, the horizon of computer-facilitated collaborative learning
has been expanded. Included among this expansion is identification of issues for general and special education students, and applications in a variety of domains, which have been suggested.
Acknowledgements

First and foremost, I would like to thank Lori Scarlatos for her guidance and support over many years. The time and patience of the rest of my committee is also greatly appreciated: Jim Cox took on a big role, Zhigang Zhu gave many useful suggestions, and Robert Haralick stepped in with constructive ideas. Much thanks to Jackie Jones for her motherly advice and support, which included huge amounts of guilt-tripping and gentle nudges. You kept me motivated and inspired; thank you. Alex Sverdlov deserves thanks for being a top-notch friend and providing insight as well as programming support. Farhan Bukhari was always in the lab to help troubleshoot and John Jannone assisted with the Vicon motion capture system and the Max/MSP patch. Louise Hainline gave guidance on experimental design. Jill Cirasella, former librarian at Brooklyn College helped track down articles. Chavie Cogan read drafts and offered suggestions and corrections.

Other Brooklyn College professors I must acknowledge: Noson Yanofsky, Yedidyah Langsam, Paula Whitlock, and Keith Harrow of the CIS department, and Noemi Halpern of the Math department.

Nancy Schneier, I can’t thank you enough for all of your help. The dissertation would not have been finished without you. Shout-out to Bruce Omori, friend and award-winning photographer, whose art gallery in Hilo, HI is but a simple rectangle. The photographs it contains, however, are an inspiration to remember to look for the beauty all around us.

Thank you to my husband, Bryan Goldstein, for prioritizing family, for encouraging adventure, and for supporting the family financially during my many years as a student. My sisters B.B. Shaulson and Yardena Mishiyev helped out by babysitting. Thanks also to my
kids: Tamara Lee, who was happy, overall, to join me in play and pictures; Nettie, for always looking for ways to help out; and Ezra, for being impossibly sweet and loving. Milton Biron, Tova Rosenbaum, Milcah Zirkin, and Fraida Itty Landy also deserve my acknowledgement.

This dissertation is dedicated to the memories of Natalie J. Landy and Bertha Biron.

My research has been supported in part by the National Science Foundation, award number CNS-0420996.
# Contents

**Abstract** iv

**Acknowledgements** vi

1. **Introduction** 1

2. **Motivation** 3

   2.1 **Education** ................................................................. 3

      2.1.1 **Practice** .......................................................... 5

      2.1.2 **Math Education** ............................................... 7

      2.1.3 **Special Education** .............................................. 8

         2.1.3.1 **Autism Spectrum Disorders (ASDs)** ................. 9

   2.2 **Social Interaction** .................................................. 12

      2.2.1 **Some Examples of Computer-Supported Social Interaction** 13

      2.2.2 **ASDs and Social Interaction** .................................. 14

   2.3 **Collaboration** ........................................................ 16

      2.3.1 **Collaborative Learning** ....................................... 16

      2.3.2 **Teaching Collaboration** ....................................... 18

   2.4 **Movement** .............................................................. 19

   2.5 **Summary** ............................................................... 22
3 Technological Background

3.1 User Interfaces .................................................. 24

3.1.1 WIMP GUIs .................................................. 24

3.1.2 Post-WIMP Interfaces ...................................... 25

3.2 Input Devices .................................................... 27

3.2.1 Multiple Input Devices ...................................... 27

3.3 Tangible User Interfaces (TUIs) ................................ 29

3.3.1 TUI Design .................................................. 30

3.3.1.1 General vs. Specific ................................... 30

3.3.1.2 Tabletop Technologies ................................. 32

3.3.1.3 Beyond Audio-Visual Feedback ..................... 33

3.3.2 TUIs as Input Devices for Children .................... 34

3.4 Virtual Environments ............................................ 36

3.4.1 Virtual Input Devices ....................................... 38

3.4.2 Virtual Output Devices .................................... 40

3.4.3 VR as an Educational Aid ................................. 41

3.4.4 A Selection of VE Applications ......................... 42

3.5 Tracking ........................................................ 43

3.5.1 Tracking Objects ........................................... 43

3.5.2 Tracking People ............................................. 49

3.5.3 Gesture Recognition ....................................... 54

3.6 Summary ........................................................ 59

4 Tangible Interfaces for Collaborative Learning Environments (TICLE) ........................................ 62

4.1 Two-Dimensional Tangible User Interfaces ............... 63

4.1.1 Related Work ................................................ 64

4.1.2 Tangram ..................................................... 66

4.1.3 Case Study 1 ................................................ 69
List of Tables

3.1 Comparison of communication techniques used in TUIs with internal processors. 45
3.2 Comparison of common identification technologies used in TUIs. . . . . . . . 48
3.3 Comparison of technologies for tracking people’s locations and/or orientation 53
3.4 Comparison of device-based gesture recognition approaches . . . . . . . . . . 56
3.5 Comparison of vision-based gesture recognition approaches . . . . . . . . . . 60
4.1 Observed behaviors for sample team . . . . . . . . . . . . . . . . . . . . . . 76
4.2 Summary of observations of the first part of the usability test . . . . . . . 78
4.3 Solved tangram: $\chi^2$ expected and observed values . . . . . . . . . . . . 80
4.4 Rankings for Mann-Whitney $U$ test . . . . . . . . . . . . . . . . . . . . . . 81
List of Figures

4.1 Tangram pieces ...................................................... 63
4.2 Some tangram shapes ................................. 64
4.3 Identifying tags ................................................. 67
4.4 Diagram of TICLE table setup ................................ 67
4.5 Screen shot of a hint given by TICLE .......... 68
4.6 TICLE tangram at the Goudreau Museum ........ 69
4.7 Feedback approval graph ...................................... 71
4.8 Hint Improvement Suggestions ................................ 72
4.9 Children using the TICLE tangram .................. 73
4.10 Number of teams that solved each puzzle .......... 77
4.11 Soma cube ......................................................... 83
4.12 Instant Insanity color arrangement ................ 87
4.13 Color arrangement ............................................... 87
4.14 Arrangement of numbers on the cubes .......... 88
4.15 Sensor diagram .................................................... 89
4.16 Arrangement of leads ............................................ 90
4.17 Photo of prototype cubes .............................. 90
4.18 Labels for algorithm ............................................. 92
5.1 Art gallery shape requiring a minimum of three guards 102
5.2 Sample art gallery floor ........................................ 104
5.3 Gallery application startup screens ........................................... 104
5.4 Being given a hint in the gallery application ................................ 105
5.5 Protected art gallery ................................................................. 106
5.6 Displayed hint does not replicate the state of the art gallery .......... 108
5.7 Classification application start screen ........................................ 111
5.8 Adult and child classifying produce .......................................... 111
5.9 In-play screenshots ................................................................. 112
5.10 Classification application, solved ............................................ 113
5.11 Wearing markers ................................................................. 117
5.12 Max/MSP patch ................................................................. 118
5.13 Vicon screenshot ................................................................. 119
5.14 Two views of the workspace .................................................. 120
5.15 Arm shadows ................................................................. 123
A.1 Instant Insanity cubes ............................................................. 134
A.2 Instant Insanity solved ........................................................... 134
A.3 Graphs of the four instant insanity cubes .................................. 135
A.4 Graph of all four instant insanity cubes combined ....................... 136
A.5 Solution subgraphs ............................................................. 136
Chapter 1. Introduction

Although a world leader in many areas, the United States lags behind many other countries in educating its children. To maximize learning, especially in the objective and technologically important areas of math and science, educators must implement techniques which are flexible enough to facilitate learning in a diverse population. This includes minorities, those with varying degrees of disability, and members of both genders.

Technology has, in recent years, been looked toward as the magic bullet for solving this educational dilemma. However, technology is clearly not a one-size-fits-all solution. Each child’s personality and learning abilities comes into play as well, with some being able to tolerate, and possibly even thrive, with one-on-one computer learning, while others come to a standstill. While working collaboratively on a single computer with a traditional interface has its drawbacks, collaborative learning enables children to assimilate larger amounts of information in a shorter period of time, so supporting this type of learning is essential.

So the question becomes, “How can we support collaborative learning of objective topics (with clear correct and incorrect answers) for learners of all backgrounds and abilities?” Yes, the answer involves the use of technology, but not in the traditional mode of interaction. Instead, we look at two different ways in which computers discreetly provide support, not demanding to be the center of attention, but rather responding to the natural manner in which people interact with other people and objects within their environment. Both systems, one using a tangible interface and one a partially immersive virtual environment, take aspects of the physical world (interpersonal communication, shared workspace, kinesthetics) and combine them with the digital world (high customizability, stored knowledge base, objectivity), to enable not just the acquisition of knowledge, but the formation of learning
strategies that generalize to future learning in a wide variety of situations, i.e., learning how to learn.

The benefits of collaborative learning systems are not limited to intellectual advancement, but have been shown to have positive effects on a child’s social confidence. As the computer system can be manipulated to be sensitive to the social needs of users, children, including those on the autism spectrum, may use these systems to learn how to function in an environment where social collaboration is important to the learning process. This may enable them to develop the social tolerance for group participation that is needed in academic, work, and social contexts.

These benefits make the use of collaborative computer learning important to teachers (both general and special education), students, parents and family (of those both with and without social issues), educational policy makers, and statewide legislators charged with investing in school support to improve performance.

The first approach used in this dissertation was different from others in that it combined tangible interaction with a non-dominating computer, guide-on-the-side, and collaborative learning. The second approach, while based on similar concepts, is different from others in that it combines a partially immersive environment with collaborative learning, kinesthetic awareness, problem solving, and a guide-on-the-side. Students use their physical location (with its unique perspective) and body movement (pointing) to interact with other students using the guide-on-the-side to arrive at solutions to various problems. With all of these elements coming together in the learning process, learning is enhanced.
Chapter 2. Motivation

In this chapter, the motivation for this research is discussed: why is supporting collaborative learning so important? This dissertation talks about the importance of education, social interaction, collaboration, and movement, in the development of a child. These are discussed both in general, as well as in the context of being computer-supported. Background for why our techniques will be useful across a broad range of learners, from those in general education to those with special needs, is also examined.

2.1 Education

A child’s informal education begins at home, most often through storytelling, when a parent or caregiver shares a story. Storytelling has been used for thousands of years to pass on a society’s values, and continues to be used to educate people in areas like medicine, law, and business.\[3\] According to Andrews et al., stories facilitate learning both directly (through speech and language) and indirectly (“by aiding in the mental construction of a sequence of events enacted for or by the learner”) through the use of story mechanisms like plots which help focus the learner’s attention.\[3\]

Although children enjoy passively listening to a story, they also want to be active participants in this pastime and, as will be discussed in more detail, a hands-on approach is an excellent way to involve children in the learning process and thereby better engage their attention. Some researchers have created computer-supported storytelling environments, such as KidPad\[37\], an extension of it using a “magic carpet”\[163\], and StoryRooms\[107\]. KidPad\[37\] allows children to draw characters and props and then use these creations to tell
a story. Adding a “magic carpet” to this environment shifts the focus from the computer by placing pressure sensor mats beneath a piece of carpet that acts as a stage. Stepping on different parts of this magic carpet—and therefore different sensor mats—navigates within the “world” that the students have drawn. StoryRooms allows children to add functionality, like light and speech, to props in their story, thereby making it more engaging by appealing to more senses.

While storytelling certainly has its benefits, when it evolves into the primary means of conveying information, problems may arise. For the most part, children today learn much as they did one hundred years ago: sitting at desks and listening as the teacher speaks. Regrettably, when taught in this fashion, many children have a hard time grasping concepts, particularly in math and science. Fortunately, in the early 20th century, Maria Montessori introduced the Montessori method, a hands-on, materials-centered approach where children interact with specially designed materials developed to appeal to, and stimulate, the senses, and slowly the approach to education began to change. Jean Piaget, considered to be one of the pioneers of child psychology, furthered Montessori’s approach with his advocacy of constructivism, which, simply put, states that people learn through their experiences. Actively doing something provides an experience, whereas passively listening to a lecture does not.

Although it has taken many years for the experiential approach to learning to find approval and become widespread, its advantages are helping to move it into the mainstream. Using objects that can be manipulated (manipulatives) in the classroom is not without its drawbacks. As students’ attention shifts from the front of the classroom to the manipulatives on their desks, it becomes harder for the teacher to address specific issues each child is having, since the teacher must divide her time among all students. Moreover, manipulative-based learning activities are often open-ended: it is up to the student to make a discovery. While this type of learning has its advantages, i.e., because no person or thing is guiding them toward a set goal, they may become side-tracked and not learn the intended lesson. To
quote Polya, if students are “left alone . . . without any help or with insufficient help, [they]
may make no progress at all.”[125] This supports evidence which suggests that a combination
of hands-on learning plus lecture may be more effective than either one alone.[17]

The following subsections deal with three educational topics that are relevant to the
research discussed in this dissertation. These include practice, math education, and special
education.

2.1.1 Practice

Regardless of whether a manipulative-based approach or lecture-based approach is used in
the classroom, it is important to reinforce the concepts with plenty of practice to ensure
that they become ingrained in the student’s mind.[17] It is widely accepted that having a
knowledgeable instructor available on hand during practice is beneficial. Educators have
come to see the role of teacher as one of facilitator, asking relevant and progressively more
specific questions to get the students to think about the problem in ways that will lead to
understanding and eventual solution.[17] The goal is to continually reduce the assistance
until the student is capable of solving the exercises unaided. This type of teaching is called
scaffolding.[1]

While human teachers are generally best for conveying new concepts, it is not always
possible for the teacher to work one-on-one with each student as they are practicing the
material. Computers are well equipped to guide, encourage, and otherwise keep students
on task until the teacher is available to work with them. While there is no evidence which
suggests that computers are better than humans as educators, they do have some benefits
that researchers are taking advantage of: they are more easily duplicated, updated, and
replaced, and, without the element of emotion, computers don’t lose patience or grow bored
as time passes, and can display a hint or offer encouragement with just as much enthusiasm

\footnote{Much like construction scaffolding, instructional scaffolding is a form of support, albeit one designed for
the learning rather than the building process. This learning support builds on Vygotsky’s Zone of Proximal
Development mentioned in §2.2 with the adult adjusting the level of help to the child’s needs, ranging from
highly specific instructions when the child is struggling, to more general help when they are doing well.}
the tenth time as the first.

For human teachers in a classroom setting, there exists the question of “to whom do they teach?” Regardless of whether they teach to the bright students, the average students, or the slower students, the needs of those in the other groups are not met. On the one hand, advanced students are denied the ability to progress, while on the other hand, slower students can not keep up with instruction ahead of their ability. Computers, in their typical one-to-one setup, do not have this limitation and can guide learning at the individual’s educational level.

Fortunately, computers, once considered adult-only tools, have slowly become accessible to even the youngest children. By making use of new computer-human interaction techniques that do not rely on good hand-eye coordination or the ability to read, computers can support learners of all ages and abilities.

One way to support learning is with Intelligent Tutoring Systems that store a representation of the student’s knowledge. They are able to solve the problems presented by the system and can give help when the student needs it. According to Sleeman and Brown, “Intelligent tutorial systems take the form of computer-based 1) problem-solving monitors, 2) coaches, 3) laboratory instructors and 4) consultants.”[160] They do all this by providing feedback and often prodding the user forward in the proper direction. A guide-on-the-side, less knowledgeable than an intelligent tutoring system, guides students to discover knowledge not by giving them step-by-step help or being able to solve problems, but rather by steering the students with relevant hints at the appropriate times. While some researchers have developed environments to replace the human teacher[21], the general view is that these systems are meant to augment them.
2.1.2 Math Education

While many children delight in storytelling\(^2\), math often has the opposite effect: many children dread it. As Papert writes, “For most people, mathematics is taught and taken as medicine.”\(^1\)[117] Yet, the skills that learning mathematics teach, like logical, organized thinking and the ability to reason, are necessary in today’s workforce.\(^1\)[133] Furthermore, of all high school subjects, it is proficiency in math which appears to have the most significant influence on the economic welfare of both the individual (better cognitive skills lead to better pay) and the country (better cognitive skills promote economic growth).\(^1\)[58]

Because of this, much attention is paid to math education by educators and education researchers. Given that Piaget\(^1\)[122] and others agree that the study of mathematics can be brought down from the abstract through the use of manipulatives, a variety of Montessori-inspired manipulatives have been introduced for math education. They include the tangram, Cuisenaire rods, base 10 blocks, and pattern blocks.\(^1\)[187] Children use these manipulatives to learn about a wide range of topics, from addition and subtraction to fractions and polynomial equations.

While manipulatives are helpful for learning concepts that deal with two-dimensional objects, they are invaluable when it comes to learning three-dimensional concepts, which are also traditionally taught using two-dimensional items like paper, textbooks, and chalkboard. Post\(^1\)[126] condemns the use of textbooks for mathematics as they contain symbols and pictures of things, but can not contain the things themselves. This shortcoming is amplified when learning 3D concepts, where much data is lost in the conversion to 2D representation (see more in §3.1.1).\(^3\)

Researchers are working to address this inadequacy. HyperGami\(^1\)[38] \(^3\)[39] attempts to help users visualize three-dimensional shapes by allowing them to print out and fold into 3D form their own customized polyhedral shapes. The researchers believe that HyperGami may

\(^2\)For its own sake, with or without enhancements.

\(^3\)Although, being able to recognize three dimensions from a drawing is important, and is necessary for such tasks as interpreting IKEA diagrams.
eventually be used to aid in the development of spatial visualization skills. VRMath aims to help elementary school children learn 3D geometry concepts and improve their spatial abilities by enabling them to build and manipulate 3D computer worlds. AquaMOOSE is a 3D environment which allows students to design 3D graphical forms by specifying mathematical equations; it is an “attempt to synergistically combine mathematics and art,” thereby appealing to a wider range of students.

These three applications (with the possible exception of HyperGami) are still using 2D media, namely the computer monitor, to convey 3D concepts. Cognitive Cubes use physical three-dimensional cubes to assess (rather than teach) users’ spatial and constructional abilities, yet it holds promise as an educational tool.

2.1.3 Special Education

Encyclopædia Britannica defines special education as:

“the education of children who differ socially, mentally, or physically from the average to such an extent that they require modifications of usual school practices. Special education serves children with emotional, behavioral, or cognitive impairments or with intellectual, hearing, vision, speech, or learning disabilities; gifted children with advanced academic abilities; and children with orthopedic or neurological impairments.”

This umbrella term covers children with a large variety of limitations, many of whom can benefit from new or unusual approaches to education. For example, given that hearing-impaired students are unable to “relate letter groupings to phonetics” but rather learn to read by recognizing word shape, and because English does not translate word-by-word into American Sign Language (ASL) and vice versa, “even the best twelve-year-old students can only read at a first or second grade level.” Aside from (or in conjunction with) using

4 These are applied ActiveCubes, which are discussed more in Chapter 3.
an ASL translator in the classroom, taking a constructivist approach allows children to learn without placing additional demands on their concentration, by not compelling them to “associate word shapes, lip movements, ASL gestures, and the meaning of the words”[152] concurrent with the learning process.

Falcão and Price[42] discuss using tangibles5 for those with learning disabilities, stating that because they often present with distractibility, poor verbal memory, poor logical reasoning, abstract thinking, and immature social and emotional skills, among other limitations, and because recommended teaching strategies include using a visual, auditory, and kinesthetic approach and cooperative learning groups, “The physicality and multisensory aspect of tangibles make them particularly suitable for children with special needs.”

One challenge of incorporating manipulatives into the curriculum for those with learning disabilities is their lack of structure. While open-ended exploration, as encouraged by manipulating physical objects, works well in general education, teachers typically prefer structured tasks for those with learning disabilities, as these children “are not capable of the self-directed learning required in the constructivist theory.”[86] Guidance can be incorporated into tangible interaction in the form of a guide-on-the-side or intelligent tutoring system, as discussed earlier and, if implemented well, may provide enough structure to balance the inherent open-endedness of the constructivist approach.

Children with autism and related disorders are also included among those who receive special education and who may derive tremendous educational benefit from non-traditional computer interfaces.

2.1.3.1 Autism Spectrum Disorders (ASDs)

According to the Encyclopædia Britannica, an ASD is “any of a group of neurobiological disorders that are characterized by deficits in social interaction and communication and by

---

5This dissertation uses the term **tangibles** to refer to computer-enhanced manipulatives.
6Being in the UK, Falcão and Price refer to learning disabilities as **learning difficulties**.
abnormalities in behaviours, interests, and activities.”

Socially, these deficits manifest as difficulties in interpreting body language and facial expressions and in understanding social norms. Intellectually, people with autism can be quite bright—Kanner’s seminal paper on autism included a case study of a child who had an I.Q. of 140—but they have difficulty with communication. Indeed, autistic author Temple Grandin, in her book *Thinking in Pictures*, says that “One of the most profound mysteries of autism has been the remarkable ability of most autistic people to excel at visual spatial skills while performing so poorly at verbal skills.” Although Howard Gardner’s theory of multiple intelligences which proposes the idea that humans possess not one kind of intelligence, but seven distinct intelligences—musical, bodily-kinesthetic, logical-mathematical, linguistic, spatial, interpersonal, and intrapersonal—explains Grandin’s mystery to an extent, it is nonetheless a genuine hurdle to be addressed.

Difficulties pertaining to learning are present as well, often in the form of difficulty staying on task, applying previously acquired knowledge in new environments, extrapolating skills, and working collaboratively. These obstacles to learning necessitate the repeated practice of skills in all variations. Tito Mukhopadhyay, in his autobiographical book *How Can I Talk if My Lips Don’t Move?: Inside My Autistic Mind*, says,

“when I am used to situations, and have labeled the objects included in that situation many times, [only then can I] label the situations and objects on my first step. And so, practice, exposure, and experience with objects and around objects matter a great deal, in order to accommodate new situations.”

Additionally, more than 90% of children on the autism spectrum suffer from sensory processing disorders, which interferes with the way that they process external stimuli.

---

7There appears to be some confusion over which of the five Pervasive Developmental Disorders listed in the DSM-IV [Autistic Disorder (autism), Rett’s Disorder, Childhood Disintegrative Disorder, Asperger’s Disorder, and Pervasive Developmental Disorder Not Otherwise Specified (PDD-NOS)] are included in the term *Autism Spectrum Disorders*, with the American Psychiatric Association’s DSM-5 Development website including all but Rett’s Disorder, while online encyclopedias like Britannica and Wikipedia leave out Childhood Disintegrative Disorder as well.

8Not taking into account those with savant syndrome, which is often associated with autism.
This may manifest as over- or under-sensitivity to stimuli of any of the senses. Both avoiding and craving the stimulation of the affected sense(s) can have an impact on learning in a number of ways, e.g., causing students to become distracted or disruptive, or having a hard time transitioning from one activity to another.

As mentioned in §2.1.1, using computers as a practice tool can be beneficial for students. This applies even more so to those with autism, as a computer’s impersonal instructions and questions mean users don’t have to worry about social concerns. In comparing autistic children’s playing of a game with human facilitation versus computer-only, Piper et al. found that much to the therapist’s chagrin, the children did better without her involvement, since they had to respond to her when she was facilitating it. Eliminating this one social interaction that otherwise would have demanded their attention, and leaving just the objective computer to enforce the rules, made playing easier. Kanner, who first described autism, refers to the autistic person’s “inability to experience wholes without full attention to the constituent parts.” Omitting one of the constituent parts—and one they found difficult at that—meant less of their attention is drawn away from the game.

Grandin believes that the predictability of computers explains, in part, why people with autism are drawn to them. They thrive on routine—even inflexibility—and what is more rote than a computer? Computers do not react to the odd behavior typically associated with autistic people, do not expect verbal responses, and can be adapted to the special needs of the child, e.g., nonverbal autistic children can interact with computers by means of gestures, or levels of visual or auditory outputs can be controlled to suit sensory requirements.

Youngblut is of the opinion that Virtual Reality (VR) is better suited to support visually oriented learners than traditional teaching media. Grandin’s suggestion that people with ASDs are primarily visual thinkers, as she herself is, implies that VR may be an excellent teaching tool for ASD students, and indeed, it has been extensively investigated as such. For
those with low-functioning autism, VR may enable users to practice real world behaviors and activities (e.g., tying shoelaces, writing the alphabet, or crossing the street safely) repeatedly, thereby increasing their independence. VR can be useful for those on the autism spectrum that are higher functioning as well, as discussed in the next section.

2.2 Social Interaction

Child defines socialization as the “process by which an individual, born with behavioral potentialities of enormously wide range, is led to develop actual behavior which is confined within a much narrower range—the range of what is customary and acceptable for him according to the standards of his group.” In other words, socialization is the process through which an individual learns to adjust to society and behave in a manner consistent with the society’s norms. For a culture to survive, the people need social experiences to learn its beliefs and practices and to pass them on. In the words of educational psychologist A.D. Pellegrini, “You don’t become socially competent via teachers telling you how to behave. You learn those skills by interacting with peers, learning what’s acceptable, what’s not acceptable.” Social interaction is, therefore, of utmost importance for both the individual and society as a whole.

Vygotsky took it a step further with his Zone of Proximal Development theory, which views all learning as a process that requires social interaction and interpersonal cooperation. His rhetorical question of, “can it be doubted that children learn speech from adults . . . or that through imitating adults . . . children develop an entire repository of skills?” can not be denied.

This implies that it is of paramount importance to support social interaction and collaboration among youngsters, and in particular, among youngsters with deficiencies in those very areas, such as those on the autism spectrum: difficulties in social interaction left unaddressed may affect not just interpersonal relationships, but also learning. Even more
troubling is that these problems may not be easy to reverse: Hoffmann and Spengler show that early social experiences can affect how genes are expressed, which in turn has a lifelong impact.[63] Thus, early social problems can create lifelong negative consequences.

The highly social and physical nature of children’s play of yesteryear, in particular sports, has unfortunately been negatively impacted by the electronic age. Combining the availability of electronic devices with parents’ relatively newfound fear of allowing their children to be left outside unattended, forms a generation of children whose most worked muscles are in their fingers. Health risks aside, the level and type of interaction are not the same as out on the field, even for those who do gather to play electronic games. Add the internet to the mix, and there is no need to gather anymore; games can be played together remotely. The limited interaction necessary to “socialize” across the internet does away with many aspects inherent in traditional social interaction, such as gestures, body language, and gaze.[98]

To counteract this decline, researchers are looking at ways to support interaction among co-located users, i.e., users that are located in the same physical space. New modes of computer-human interaction pave the way for computer-supported social interaction, and in fact, according to Hornecker and Buur, “the support of social interaction and collaboration might be the most important and domain-independent feature of tangible interaction.”[64]

### 2.2.1 Some Examples of Computer-Supported Social Interaction

Creighton[31] presents ‘jogo’, a tabletop music game intended to encourage physical activity and social interaction while children explore creating music. Played on a circular table to encourage interaction—there are no “sides”—brightly colored balls are placed in various locations on the table to produce music. Creighton believed that jogo’s “form, sound, rhythm, and creativity” would motivate children to play together and, upon observation, noted this was indeed the case.

With Playground Architect, Hendrix et al.[60] aimed specifically to encourage shy children to take on a leadership role and gain social confidence. For shy children,
fear of failure is often the cause of their inhibition to initiate social contact, but with opportunities for social success among peers, they may be able to develop social competence. Playground Architect, a multiplayer tabletop game, involves laying out the various components of a playground according to a set of rules, and was designed to give shy children a “success opportunity” to help bolster their social confidence. The role of architect—straightforward but essential to the successful completion of the task—was assigned to the shy child, where only they were given the layout specifications, facilitating their moving into a leadership role by requiring them to pass on instructions and make decisions. Upon evaluation, the researchers found that the shy children enjoyed being in charge and, according to their teachers, behaved in a more outgoing fashion while in the role.

Other researchers have concentrated on supporting social interaction across a network, especially useful for the times and conditions when getting together with others is not possible, such as living in a remote area or being isolated for health reasons. PlayPals allows girls to play dolls with a friend remotely. Each girl has two dolls, one theirs to play with, and the other representing the doll of their remote friend. Each pair of dolls is synchronized, so when one child moves her doll’s hand, the matched remote doll’s hand moves as well. Various accessories allow the playmates to communicate and share multimedia content.

2.2.2 ASDs and Social Interaction

Social interaction is important even for those who find it difficult, like those on the autism spectrum. Grandin discussed meeting Tito Mukhopadhyay, another person with autism, albeit nonverbal and altogether much lower functioning than herself, whose mother had taught him to type. When she asked him what it was like before he could type, his answer was a single word: “emptiness.” Although Grandin sees the benefits of computers in
providing some level of social interaction for those with ASDs, it may be more constructive to use computers not as an escape, but as a means to teach improved social interaction.

While researchers are making progress with social interaction support for those that are high functioning, doing the same for lower functioning people on the spectrum is more complicated. Keay-Bright and Howarth, in an attempt to fill this need, aim to engage the user through “action, rather than cognition,” arousing curiosity and making interaction irresistible. This approach has been successful in encouraging social interaction in low-functioning people with autism, with the following behaviors displayed: sharing, a greater variety of sounds being produced, and decreased echolalia.

Asperger’s Syndrome (AS), also called High-Functioning Autism (HFA), is characterized primarily by significantly impaired social interaction, and people with AS may suffer from social isolation because of their difficulties in making and keeping friends. Social skills training can teach them strategies to navigate social situations, but is most successful “either in situ or in role-play situations where users can explore different outcomes resulting from their social behavior.” Since social gaffes made in practicing these skills in real-world situations may leave the person feeling anxious or embarrassed, which can hinder progress, role play within a computer—or virtual—environment can instead address social issues for those across the spectrum by providing a safe life-like environment in which to practice social skills. Within such an environment, all variables can be controlled to maintain uniformity or modified as needed, while minimizing the necessary social interactions reduces associated anxiety. Parsons and Mitchell emphasize that “the idea is not for the [virtual environment] to minimize social interactions per se, but rather, to allow the safe and non-threatening practice of particular skills in an educational setting.” This enables those with ASDs to become comfortable with these skills, at which point they can be practiced in a real-world

---

10 Computers allow them to avoid problematic aspects of face-to-face contact such as having to make eye contact, and awkward gestures are not visible in typewritten messages.
11 A phenomenon often associated with autism, Merriam-Webster Online Dictionary defines echolalia as “the often pathological repetition of what is said by other people as if echoing them.”
12 Good for repeated practice and reducing distractions.
environment. Virtual reality is further discussed, in depth, in §3.4 and a selection of VR applications for those with ASDs is discussed in §5.1.

Computers can help train improved social interaction among peers, too. One example is the Augmented Knights Castle (AKC), a Playmobil® knights’ castle digitally augmented with sound[62]. Placing various figures in active areas cause them to “speak.” While all children can enjoy playing with the AKC, the researchers adapted it for children with autism by allowing the children to program where, when, and what the figures said. Farr et al. found that with this configurability, “less solitary play and more cooperative play occurred.”[44] They believe that giving them greater control allowed for more socially oriented behavior because of the broader range of interaction styles available.

2.3 Collaboration

Collaboration is a form of social interaction with the added element of a shared goal. We move on to a discussion of its use in learning, i.e., learning as a collaborative effort, and discuss teaching collaboration to those with social difficulties impacting their ability to collaborate.

2.3.1 Collaborative Learning

Collaboration can be especially effective in an educational environment. Collaborative learning is defined as groups of students performing at a variety of levels working together to achieve a common academic goal.[52]

Collaborative learning requires those in the group to be responsible for other students’ learning in addition to their own. To be successful, students must join together in analyzing and thinking through a problem, exposing the collaborators to multiple points of view. Research shows that children may be more motivated to contribute when working in groups[17], and are able to solve more problems that way.[68] Moreover, the discussion
that takes place may help promote critical thinking.\textsuperscript{13} Within creative and exploratory environments, too, collaboration may support learning, as differences of opinion will require additional exploration to resolve.\textsuperscript{41}

It is important to note, that although students within a group are at various performance levels, these levels should not vary too much for it to be considered collaborative, as collaboration implies that all participants share knowledge and contribute to the goal. Furthermore, current research is exploring ways to enable collaborators\textsuperscript{14} to be aware of each other’s mental and emotional states so that they can better regulate their learning, leading to more effective collaboration.\textsuperscript{15,76}

Working on a paper-based assignment as a group increases the likelihood that one student will do the bulk of the work. For the more advanced in the group, it can be a matter of getting it done quickly without waiting for input from the rest of the group, while for the slower students, it is both easier and quicker to allow others to complete the assignment and share it.

Computer-supported collaborative learning (CSCL) makes use of technology in a variety of ways to support this type of learning. CSCL has been shown to best both individual and competitive learning when collaborators are co-located.\textsuperscript{16,53}

The use of computers in supporting team assignments can be successful if “a learning environment that gives all of the children equal access to the data and equal opportunities to manipulate that data” is provided.\textsuperscript{142} In other words, true collaboration can only be achieved when all members of the team can explore simultaneously, no turn-taking required. While this is difficult to accomplish with a standard graphical interaction system, using manipulatives is perfectly suited to group learning activities: the manipulatives can be used

\textsuperscript{13}Critical thinking was defined by Scriven and Paul for the National Council for Excellence in Critical Thinking Instruction, as “the intellectually disciplined process of actively and skillfully conceptualizing, applying, analyzing, synthesizing, and/or evaluating information gathered from, or generated by, observation, experience, reflection, reasoning, or communication, as a guide to belief and action.”\textsuperscript{32}

\textsuperscript{14}This current research focuses on networked learning groups.

\textsuperscript{15}This is discussed in some more detail in the next subsection.

\textsuperscript{16}For the most part, we don’t discuss non-co-located collaborative learning, which must be supported by technology.
to represent data, which may then be easily manipulated by multiple people equally.

Students working together in a hands-on educational environment benefit all-around: manipulative objects pull all of the students’ attention toward the physical object, and may often require the assistance of an entire group to solve a problem. This translates to no student taking over and no student abstaining or becoming lost. Incorporating the support of a computer adds the benefits of CSCL.

### 2.3.2 Teaching Collaboration

Being able to collaborate effectively is an important enough skill that when it is lacking, it should be taught. Computers may be employed to support the learning of the skill of collaboration: for a particular action to be done, all users must do it together (or at least agree), with the computer enforcing this rule. Called Enforced Collaboration (EC), it has mostly been explored for those on the autism spectrum, who often need the extra help in learning to work collaboratively.\[11, 48\]

Shared Interfaces to Develop Effective Social Skills (SIDES) is a tabletop game designed to help high-functioning autism spectrum adolescents learn to play fairly and work cooperatively.\[124\] The game involves building a path to allow a frog to travel from one lily pad to another, eating insects to gather points along the way. Each of the four players receives a selection of the (virtual) pieces to be used to construct the path. Being able to distinguish among each user’s touch enables the computer to enforce turn-taking and prevent one person from “stealing” another’s pieces. Once the path is completed, all players must vote unanimously to approve it. In this way, players practice socially accepted collaboration behaviors and have them reinforced.

StoryTable uses enforced collaboration “to facilitate collaboration and positive social interaction” in children on the autism spectrum in a storytelling environment.\[48\] Participants are required to perform joint actions in order to operate some of the interface’s

\[17\] Having players sit around a table places them in a face-to-face social configuration.
functions. The researchers observed a number of positive effects on ASD children who used StoryTable including a subsequent increase in initiation of social interaction with peers, an increased level of collaboration afterward, and a relative decrease in frequency of autistic behaviors while using StoryTable.

Assisting in collaboration skill development will benefit all children, including those not on the spectrum. As the level of learning in collaborative situations is dependent on the quality of the interaction as well as on the feeling of being a group, Järvelä et al.[76] focus on supporting socially shared regulation of learning (SSRL), where collaborators are given a peek into the thoughts and feelings of their teammates. This imparts a sense of community, while simultaneously affording them the ability to regulate the group’s collective learning. The regulation of shared learning is accomplished by employing a variety of cognitive, metacognitive, behavioral, and motivational techniques, which boosts their collaborative performance.

2.4 Movement

“Movement is at the very center of young children’s lives,” writes physical educator and researcher David L. Gallahue.[49] Not only is movement important to a person’s physical health, but also for their intellectual, emotional, and social well-being. While the previously discussed jogo[31] (§2.2.1), Playground Architect[60] (§2.2.1), and AKC[62] (§2.2.2) all support social interaction, they are missing the element of movement. Whether taking place on the playground for young children or as part of a team sport for older children, play with active movement has always been an important outlet for youngsters—especially boys—to let off steam and socialize. James H. Humphrey explains that sports in particular are, by their very nature, socially-oriented. To be successful in sports, children must develop interpersonal skills such as learning to work together for the benefit of the group, accepting and respecting the rules of the game, thinking and planning cooperatively, and learning to
win and lose gracefully. [66]

Researchers are therefore investigating ways to use computers to support collaboration in active settings. Soler-Adillon et al.[161] have designed the Interactive Slide, a multi-person inflatable slide onto which is projected a game. Players must work together, climbing up and sliding down repeatedly as the game demands.

PingPongPlus[73] supports movement by computationally enhancing a Ping Pong table. Various play modes encourage various types of interaction. In the water ripple mode, an image of rippling water flows from the spot where the ball hits the table. Rather than playing competitively, players explored making interference wave patterns on the table. Another mode supports collaborative painting on the table and yet another mode encourages the players to work together to keep the ball in play.

Bekker et al. created the Swinxsbee, a Frisbee-like device used with a screen-free game controller designed to encourage group physical activity.[13] In their tests, they found that the shared use of the Swinxsbee led to higher levels of social behavior. However, physical activity that was too intense was found to diminish social interaction, as participants must focus more on playing the game.

While most people think of being co-located a requirement of playing sports, Mueller et al.[109] have come up with a way to support sports over a distance. The game they developed, Breakout for Two, supports soccer-like game play across a network, with players kicking the ball toward a life-size audio-video projection of their opponent. The researchers believe this setup “facilitates the social interaction and encourages conversations.” In a comparison between physically intense kick-the-ball-hard and keyboard versions of the game (where a virtual ball was hit by a virtual foot), players felt they bonded with the other player considerably more when playing the active version.

In addition to supporting the social aspect of movement, computers can provide support for using movement as an academic learning tool. As a person physically moves about in the world, their perspective changes, and with it, their perception of the world. Much
as *HyperGami*\(^{38, 39}\) (mentioned in §2.1.2) enables users to analyze three-dimensional shapes by physically turning them, thereby allowing users to view the shapes from different perspectives, a person’s own movement enables them to better analyze their environment, the objects within it, and the relationship among those objects. Thus, movement may help develop a person’s analytical skills, and computers can be employed to encourage this development.

Moreover, evidence that some people display a preference for learning through the kinesthetic sense\(^{102}\) has propelled educators to devise curricula which integrate physical education, math, science, and computers. Such curricula are believed to help children become more involved in the learning, attain higher-order thinking skills, and work collaboratively.\(^{59}\)

Westreich\(^{192}\) found that students taught math concepts through creative movement (dance), scored better than those in a control group. Administering a delayed post-test strengthened their findings, showing that the creative movement approach improves retention. Westreich also noticed some unexpected benefits in the experimental group: students developed increased bodily control, which decreased hyperactive behavior in those prone to it, and teachers’ attitudes toward difficult students improved when they saw their reactions to the new strategies.\(^{192}\) As experiments have shown that “autistic children learn through manipulation and position cues rather than through normal perceptual processes”\(^{115}\), kinesthetic learning experiences may address their specific needs.

Price and Rogers’s Hunting of the Snark game has children trying to discover as much as possible about the Snark, an elusive virtual creature.\(^ {128}\) They do so by walking in its cave, feeding it, and “flying” with it. Although the activities are done in pairs, their primary purpose is to provide a kinesthetic learning experience. Zhang et al. also created a computationally enhanced game to teach kids about energy and the environment while having fun jumping and dancing around.\(^ {200}\) Playing in pairs, students must use a combination of fossil fuels, renewable energy, and kinetic energy to power their object
toward a goal first, but the limited fossil fuels available in the game forced the players to create their own energy by jumping up and down to the sound of a drum roll.

2.5 Summary

To summarize, education, social interaction, collaboration, and movement are all interrelated. Education is a form of social interaction, and social interaction is a form of education. Education typically requires a teacher interacting with students, while at the same time, interaction between an elder and youngster teaches the youngster about interaction. Collaboration is an important tool in education, which benefits both strong and weak students by encouraging a sharing of ideas. It is just as important in social situations, teaching children to strive for the benefit of the group, not just the self. Movement, too, can be used as a means to educate, and is an important mode of social interaction and collaboration.

As we have seen, computers can support all of these, both individually and in various combinations, and is not restricted to users of particular abilities. Computers are capable of transcending almost all neurological, learning, and physical disabilities, to allow people to be educated according to their needs.

We will next give some background information on the technologies supporting the wide range of learning tools discussed in this chapter.
Chapter 3. Technological Background

We now delve into some technological background for the two environments in which we aim to support collaborative learning. The root of this research is user interfaces, and I begin with a discussion of that topic. As my work takes two distinct paths, I proceed with background for the first segment, beginning with a review of input devices and building up to tangible user interfaces. I then move on to review literature relating to virtual reality, the second segment. I conclude with a merged discussion of tracking. The following diagram shows this breakdown, with lower topics building on those above them. The section of the topic is in parentheses.
3.1 User Interfaces

According to Merriam-Webster Online Dictionary (www.m-w.com), the word interface means “the place at which independent and often unrelated systems meet and act on or communicate with each other.” In our case, the independent and unrelated systems are the natural world in which we live, and the virtual; the interface is where we bind these systems into one, allowing us to slip back and forth between the two realms.

3.1.1 WIMP GUIs

WIMP GUIs, which stands for Graphical User Interfaces based on Windows, Icons, Menus, and a Pointing device, have been the predominant user interface since they were introduced by Apple’s Macintosh, and popularized by Microsoft’s Windows for the PC. GUIs made possible WYSIWYG (What You See Is What You Get) word processors, that display on-screen exactly what will appear when printed. This new graphical representation was a significant improvement over the command line, and enabled young children, even those unable to read or write yet, to use computers by “pointing-and-clicking.” With touchscreen hardware, even toddlers can use a GUI.

Despite the many positive aspects of GUIs, they also have some drawbacks: too many screen-cluttering widgets can be confusing; all the layers of “point-and-click” mean that users are spending too much time manipulating the interface, not the application; and using a mouse and keyboard is associated with repetitive stress injuries.\[176\] Plus, although GUIs (for the most part) work well for applications that are inherently two-dimensional, such as spreadsheets and word processors, when it comes to anything even remotely three-dimensional, such as CAD or game-playing, GUIs’ shortcomings become more clearly evident. Critical information is lost when three-dimensional objects are mapped onto a two-dimensional screen, making the rotation and translation of 3D objects even more awkward. A child playing with a jigsaw puzzle is directly manipulating the puzzle pieces:
moving, turning, attaching. Playing with the equivalent puzzle on a computer, with its
text}pieces displayed on the monitor yet manipulated with a mouse (or even a touchscreen!) is
quite a different matter. Xie et al. [194] observed “that some children had difficulty rotating
GUI based pieces.” Rotating jigsaw pieces in a GUI\(^1\) is such a challenge, that at least
two online jigsaw puzzle games this researcher came across did not require the pieces be
rotated; one had all the pieces displayed in the correct orientation, while the other corrected
the orientation of each piece as the user clicked on it, thus removing a large part of the
challenge. Another difference: pieces that partially overlap on the computer monitor don’t
have that telltale bump indicating another piece is hidden beneath.

Ishii [70] says that, “The GUI, tied down as it is to the screen, windows, mouse and
keyboard, is utterly divorced from the way interaction takes place in the physical world.”
One example of this is the idea of “dragging”: in real life, we don’t drag lightweight items
such as puzzle pieces from point A to point B; we pick them up and deposit them elsewhere.
Also, as Sharlin et al. [157] state, “When we work with physical objects, we perceive both our
fingers and the objects they handle in the same time and space.” We don’t have this same
inherent awareness when working with a GUI. A user must translate her/his request into a
format that the interface can handle, and accept a response that is formatted by the computer
to conform to the output device. Ishii [70] refers to this as a “spatial discontinuity between
[the input device and output device]” that’s compounded by a “multimodal inconsistency,
as touch is the main input while vision is the only output.”

### 3.1.2 Post-WIMP Interfaces

As computer-human interaction continues to evolve, it becomes clear that what we want is
almost what we started with. In the beginning, there was no interface because there was no
interaction, and this seems to be what people want to return to: no [visible] interface and no

\(^1\)Although multi-touch capability does simplify the rotation of graphical objects, and is becoming
widespread particularly on small computing devices, we omit them from the discussion for the time being
as they are not WIMP GUIs.
Van Dam defines post-WIMP interfaces as “containing at least one interaction technique not dependent on classical 2D widgets such as menus and icons.” These may include tangible user interfaces, virtual or augmented reality, gesture recognition, and speech recognition, but will ultimately “involve all senses in parallel, natural language communication and multiple users.” Van Dam further speculated that eventually, the computer interface may “metamorphose into a total sensory environment.”

A voice interface is a computer interface that takes voice commands as input. Although the voice recognition aspect of such an interface has mostly been surmounted with applications like Dragon Naturally Speaking, understanding typical spoken language is not an easy task for a computer: humans speak ambiguously, use unclear references, and structure their speech in a variety of ways. According to Lee, “gestures often identify underlying reasoning processes that the speaker did not or could not articulate providing a complementary data source for interpreting a set of utterances. Thus, gesture and speech go hand-in-hand in daily human-to-human communication, and it would be appropriate for any interactive system that attempts to provide a similar level of fluidity to be designed with that in mind.”

Bolt’s Put-That-There system implemented a speech recognition system that used gesture to disambiguate what a user means when using pronouns. As Bolt says, “Because voice can be augmented with simultaneous pointing, the free usage of pronouns becomes possible.” Children frequently like to ask questions like, “What’s that?” Taken together with what the child is pointing at, the question can be understood. Carbini et al. implemented a primarily gestural interface that was able to use voice selection, but the details were unclear. To be discussed in detail in later sections are: Gesture Interfaces, including the wearable interface SixthSense and Tangible User Interfaces, which track object manipulation. In the following section we will discuss the gadgets used to direct and give data to the computer, collectively known as input devices.
3.2 Input Devices

Input devices come in many different varieties of camera, microphone, keyboard, scanner, and mouse, among others. The average person typically interacts with their computer primarily through the keyboard and mouse. Because the keyboard is a relic from command-based interfaces, it will not be discussed. Developed by Engelbart, the mouse is a pointing device that moves a cursor on the screen. It allows for object selection, navigation, and command issuance, including typing with a virtual keyboard. The mouse has become so ubiquitous, we almost wouldn’t know how to use a computer without one. However, as with all other technology, there is room for improvement.

3.2.1 Multiple Input Devices

Since each computer has just a single mouse, the user is limited to being able to control one item at a time. At times, it may be better to control two or more items (or aspects of those items) at once. Although some of these uses may involve individual users, the true benefit of multiple input devices becomes apparent in a collaborative application.

In older-style mouse-driven applications, only one person can have the mouse at a given time. The one with the mouse is the one with control, and therefore gets to make the decisions. In experiments with two children and a single mouse on a computer, Inkpen et al. found that there appeared to be a direct link between boredom and not having control of the mouse. The majority of off-task behaviors were found to occur during the time that the child did not have control of the mouse.

Adding a second mouse which each child could control independently (with separate cursors) yielded mixed results. According to Inkpen et al., there were three main benefits: children were more engaged, they were more active, and they significantly preferred the multiple mouse setup. However, there are drawbacks as well: keeping track of which mouse pointer belongs to whom becomes increasingly confusing as more users are added.

\(^2\)For many disabled people, the microphone may be the primary input device.
and may even require individuals to physically point at their cursor or mouse pointer so others know which one is theirs.\textsuperscript{[36]} Also, giving each child control may in some cases cause non-collaboration, by allowing the children to work independently, on separate sections of the application\textsuperscript{[3]}

Triangles\textsuperscript{[54]} and Navigational Blocks\textsuperscript{[22]} are two systems that use multiple (non-mouse) input devices. Triangles\textsuperscript{[54]} uses physical triangular pieces to allow users to arrange presentations by manipulating the triangles. When two triangles are connected or disconnected, they trigger related events. Cinderella 2000 is an application using this system: seven triangles, each depicting a person or object from the Cinderella story, can be put together in numerous configurations to recount different parts of the story. For example, a triangle with an image of Cinderella’s stepmother can be snapped together with another tile representing their house. This would cause the stepmother’s voice to be heard calling, “Cinderella! Cinderella! Oh, where is that girl?” Although not required, multiple children can use this system simultaneously, each connecting or disconnecting triangles as they choose. This early system is a step up from a single mouse for multiple users, but it still has its drawbacks, such as a single parent triangle (a single focal point), through which all others must be connected.

Navigational Blocks\textsuperscript{[22]} similarly uses objects—cubes, in this case—to affect a presentation based on their orientation and connection to other pieces. Depending on which faces of the cubes are upturned, different information is displayed. Cubes can be connected to each other if they are related to narrow the data. While this system does not use a parent block, only one or two blocks can be used in conjunction with each other, which limits its usage for collaborative applications.

Multi-touch displays and users’ bodies can also act as input devices, with the latter being discussed in \textsuperscript{[3.4.1]} \textsuperscript{[3.5.3]} and Chapter \textsuperscript{[5]}. Multi-touch displays allow multiple inputs simultaneously, so a user’s fingers are understood as separate inputs. For an individual

\textsuperscript{3}Enforced Collaboration, where an action can only be performed if everyone agrees, can force children to work together, but has its own drawbacks. See \textsuperscript{[2.3.2]} for a more detailed discussion.
user, this allows the user to motion concepts like “zoom in” and “zoom out” by moving the fingers farther apart or closer together respectively, or “rotate” by rotating the fingers in place. Large scale multi-touch displays allow multiple users to interact with the display at a given time, both on collaborative and individual actions, but often recognize only a limited number of gestures. Some other issues associated with these tables are allowing fingers to be placed on the table in any order, and differentiating between like gestures.\[28\]

Tangible User Interfaces, discussed in the next section, act not only as input devices, but as the user interface itself.

### 3.3 Tangible User Interfaces (TUIs)

With Ishii and Ullmer’s\[72\] declaration that “We live between two realms: our physical environment and cyberspace. Despite our dual citizenship, the absence of seamless couplings between these parallel existences leaves a great divide between the worlds of bits and atoms,” came an explosion of tangible user interfaces (TUIs), a term they coined to refer to a new type of human-computer interaction technique that “will augment the real physical world by coupling digital information to everyday physical objects and environments,”\[72\] thereby making “digital information directly manipulable with our hands, and perceptible through our peripheral senses by physically embodying it.”\[70\] TUIs take advantage of the physical properties of objects as well as our natural manner of interacting with tangible objects: as spatial beings, we exist in a three-dimensional world and, to revisit Sharlin et al.’s quote, “perceive both our fingers and the objects they handle in the same time and space.”\[157\]

Thus, TUIs, which don’t require conversion to a two-dimensional interface, and allow us to focus on a single place—our hands manipulating the objects, and inherently providing haptic feedback\[4\]—can be more effective. Tangible objects can also be acted on without requiring immediate feedback from the computer, unlike using the mouse, for instance, whose visual position must be updated to allow effective accomplishment of tasks.

\[4\]The feeling we receive from physically holding and manipulating the objects.
Another advantage of TUIs is the multiple access points (i.e., multiple ways to access the application via the various components of the TUI system) they provide, which intrinsically promotes collaboration by distributing control and preventing one individual from taking control, as is so easily done when using a WIMP GUI. These may make it easier for shy people to join in, as discussed in §2.2.

In the coming sections, we first examine TUI design and then explore their use as input devices for children.

3.3.1 TUI Design

Unlike the components of a graphical user interface, physical objects do not disappear when the computer is shut down, and do not require a power source to exist. Because of the disconnect between the physical artifacts and the computer, additional thought must be given to the design of the tangible components to balance their usability and physical properties with computer requirements. The tracking and recognizing of tangible components is discussed in §3.5.

3.3.1.1 General vs. Specific

Norman[113] asks the question, “When you first see something you have never seen before, how do you know what to do?” and responds that “the appearance of the device could provide the critical clues required for its proper operation,” and that’s the aim of a TUI designer as well.

The literature is replete with advocates for specialized interfaces as well as advocates for general interfaces; each has its own pros and cons. On the one hand, “in the physical world, objects and their function are usually unified and inseparable, making the effect of their manipulation intuitive and easy to anticipate.”[157] With specialized TUIs, we are already aware of each item’s affordances, defined to be “the perceived and actual properties of the thing, primarily those fundamental properties that determine just how the thing could
possibly be used.”[112] This makes interfaces based on specialized artifacts more natural to
use and easier to learn. However, it is interesting to note, that since a TUI object can not,
in fact, be used exactly as it would be in the real world, it may cause confusion when it does
not act as expected. Falcão et al.[43] observed a child becoming sidetracked by a flashlight
that “worked” even though it was off.

On the other hand, there is the need for generalizations. It is not practical to store the
physical components necessary for a multitude of programs, each using specialized TUIs.
These special purpose systems also “do not scale to support many different applications, as
is common with graphical user interfaces.”[130] Although generalized interfaces have a less
natural feel, the interface remains the same regardless of the application that implements it.
In that way, the learning curve may be reduced across TUI applications, provided, of course,
that the same implementation is used.

Levy[95], in talking about some of the applications that were written for the Macintosh,
says that those that didn’t conform to the graphical specifications set forth by the
programmers, ended up being passed over by consumers for those that did conform.
Having keyboard combinations do the same thing for all applications, regardless of who
programmed each individual application, decreases the time it takes a user to learn how to
use a program. Users can often use a program without ever opening a manual. They also
don’t have to memorize a whole new set of keyboard combinations for each application that
they use. This ensures that they won’t inadvertently press some key that does action A1 in
application B1, but action A2 in application B2; e.g., delete text rather than copy it.

In contrast, current TUIs have no standard for interaction; knowing how to use one
system does not in any way lessen the time required to learn another. While this may make
sense—knowing how to use a can opener does not impact one’s knowledge of how to use
a drill[5] for instance—compatibility and standardization are possible even across dissimilar
objects.

5Usage of a GUI can opener and GUI drill may be quite similar.
Although this design debate has not been resolved, with most systems being highly specific, and some being more general, there are some researchers that are trying to find a middle ground. Rekimoto et al., for example, introduced DataTiles\textsuperscript{130} which use physical artifacts but maintain flexibility by defining them graphically.

### 3.3.1.2 Tabletop Technologies

Tables, by their very design, are intended for groups to gather around, whether it be for eating, working, or playing. Though the shape of the table may at times be dictated by the technology being used, it is often a consideration when designing a tabletop TUI. Creighton’s jogo\textsuperscript{31} and Jordà et al.’s reacTable*\textsuperscript{78, 79, 80}, both collaborative musical instruments, are designed as circular tables to encourage social interaction by making it approachable from all sides and allowing groups to form. In contrast, the sharp corners of a square or rectangular table keep people apart, which may be desirable for applications where users compete or otherwise must maintain their own workspace.\textsuperscript{154}

A table need not be used merely as a platform on which to place tangible objects, it is often designed as part of the system. Rather than displaying supplementary information on a separate display screen, and in the process dividing the users’ attention between objects in their hand and said display\textsuperscript{120, 127}, the tabletop often serves the dual purposes of display and workspace. For the display, ceiling-mounted projectors are often used, but are subject to occlusion, especially by users’ arms.

The DiamondTouch table\textsuperscript{36} is a tabletop technology that allows multiple users to touch the surface simultaneously, and can differentiate among users by having their touch complete a circuit going through the table and their body. Distinguishing among users may be useful in an educational setting and can be used to provide individualized feedback\textsuperscript{124}, and requiring users to remain seated at the table has the added benefit of requiring students to control their behavior as well as disallowing assistance from observers. Typical objects that people
might put on a tabletop, like a mug or paper would not affect the workings of this table \[6\] and special ones can be designed to function in conjunction with it \[36\]. Some applications that use the DiamondTouch table are Piper et al.’s SIDES \[124\] and Gal et al.’s StoryTable \[48\].

The DiamondTouch table is not without its shortcomings: the overhead projector subjects the display to occlusion, particularly by arms; requires caution not to bump the table thereby misaligning the projected image; may present health hazards for some; and may be difficult for those with short arms to use, as they may not be able to reach across the table while seated.

Another out-of-the-box tabletop solution is Serious Toys\(\text{®}\)’s TagTiles \[168\], an appealing tool for tabletop TUI development, leaving the engineering aspects out of the equation and allowing designers to concentrate on the design of the TUI. Unlike the old-fashioned board game, TagTiles can adjust to the level of the users and provide help as needed. One fascinating educational game built on TagTiles, by Zhang et al. \[200\], helps users learn about different kinds of energy and their effect on the environment. It is previously discussed in §\[2.4\].

TagTiles and newer multi-touch tables \[7\] that are becoming popular for supporting collaborative activities, are themselves the display; therefore they are not subject to occlusion, and some can even recognize tagged items placed on them. \[26\]

3.3.1.3 Beyond Audio-Visual Feedback

Despite the tangibility of TUIs adding an extra dimension to computer-human interaction, audio and visual are still the dominant output modes, a fact which researchers are working to change.

Just as a GUI is capable of moving icons, etc. on its own (think re-ordering files alphabetically), TUIs can at times be more useful if given this capability, e.g., Senseboard \[73\].

---

6But Scott et al. \[154\] suggest “a mechanism that recognizes the placement of an object [so as to] not display relevant information in the physical space occupied by the item.”

7Discussed in §\[3.2.1\]
is a TUI that allows users to arrange information on a grid, such as for scheduling, yet if able to move the tangible pieces, it would be capable of suggesting arrangements as well.

Modifying the digital representation of a tangible object would accomplish nothing if the physical object would have no means of being modified; tangible objects exist in 3D space, and have a space, texture, and position, which can not be changed as easily as modifying a pixel on the screen. Navigational Blocks \cite{22}, although not truly moving, uses embedded electromagnets to give users haptic feedback: if two objects can be grouped together, the electromagnets turn on, causing the objects to attract each other; if there is a conflict, they repel each other; and if no relationship exists, the magnets are turned off. Pangaro et al.'s Actuated Workbench \cite{116} uses electromagnets to move magnetic pucks across a tabletop\footnote{The authors describe how they could rotate their pucks if each contained two magnets, but had not implemented it as of the cited paper.}. Rosenfeld et al.'s Planar Manipulator Display \cite{134} implements bidirectionally movable\footnote{May be moved by the computer as well as the user.} physical objects as mini two-wheeled remote-control cars, that may rotate as well as move; and Bonanni et al.'s PlayPals \cite{16} arms are moved by the computer according to how a remote doll’s arms are moved. Surflex \cite{30} is “a programmable surface that can take new shapes [using] shape-memory materials,” which, with enough pixels, can be used to create new objects as necessary, rather than just moving or rotating a part of an object.

Further means for allowing tangible objects to self-modify are discussed in the section on virtual environments (\S3.4).

\subsection{3.3.2 TUIs as Input Devices for Children}

Although technology is often created for serious purposes, it tends to make the move towards more playful usages as well, so it’s not surprising that the computer, which was originally invented for computation would move into the children’s entertainment arena. Having age-appropriate content is not sufficient; without suitable input devices, it remains much like TV, encouraging passivity. Johnson et al. \cite{77} configured stuffed toys to be input devices, as
controls for digital representations of themselves, e.g., shaking the toy in a running motion causes the onscreen avatar to run. Bruikman et al.\cite{Bruikman2006} uses a similar approach in their Lali, a toy for toddlers: “Touching physical Lali’s nose causes the virtual Lali to sneeze. Hugging Lali causes blushing, and above a certain pressure threshold hearts would radiate from the body. Squeezing the paws makes the arms rotate upward independently. Furthermore, touching a sensor in physical Lali’s navel alternately causes the virtual Lali to burp or giggle. Virtual Lali’s tail was animated to wag automatically every 25 seconds.”

While adults require a high level of control and expect their input devices to give it to them, children, and especially the youngest ones, lack the coordination and even inclination for such a fine level of control (i.e., they don’t want to be forced to move a mouse to such and such pixel on the screen and manipulate it). Since Johnson et al.\cite{Johnson2005} believe that, “The interface should use context to disambiguate the input rather than forcing the user to do so,” their plush toy input devices try “to understand the intentions of the user’s input in the current context of the virtual environment and tries to help them achieve it. For example, if the user is clearly heading [to a specific location], the [computer] should realize this and help navigate there rather than forcing the user to make fine control adjustments.”

Tonka Workshop\cite{TonkaWorkshop} is a platform with all sorts of play tools that attaches to the top of a keyboard and are used to complete on-screen tasks. On the amazon.com webpage for this toy, is the description, “by manipulating the playset’s screwdriver, hammer, drill and saw, kids can build onscreen robots, toy houses and spaceships.” According to Sharlin et al.\cite{Sharlin2009}, “The tools offer simple, spring-based passive haptic feedback when actions are performed with them, and the Workshop reinforces this feedback with additional audio and visual feedback.”

Tangible user interfaces are becoming almost commonplace in educational settings, and will be discussed in some detail in the next chapter. Now, I’ll change focus slightly to give background on another type of post-graphical interface.

\footnote{Unfortunately, this toy has been discontinued by the manufacturer.}
3.4 Virtual Environments

Virtual Environments (VE), often called Virtual Reality (VR), are synthetic environments in which the user is immersed to varying degrees, at times completely immersed and being unable to see the real world. One of the first such systems was Morton Heilig’s *Sensorama*, a mechanical device, which predates personal computers. Rheingold described the *Sensorama*:

“I put my hands on the handlebars and rested my face against a viewer that looked like a pair of binoculars with a padded faceplate. Right below the eyepiece was a small grill, near my nose, where the odors would have been pumped in and out of smelling range. Other grills to either side of my face emitted unscented breezes at appropriate times. Small speakers were positioned on either side of my ears.”

The experience included a motorcycle ride through the streets of 1950s Brooklyn and a dune buggy ride. Although Rheingold describes hearing the automobile engine, seeing the expanse of sand dune, and feeling the seat lurch, the experience was not Virtual Reality in the usual sense, as it was not interactive. He had no way of steering the virtual vehicles or changing the course of his ride; he was merely a passenger in the driver’s seat.

According to Witmer et al., “The effectiveness of virtual environments (VEs) has often been linked to the sense of presence (“being there”) reported by users of those VEs,” where “Presence is defined as the subjective experience of being in one place or environment, even when one is physically situated in another.” In VR, this would be experiencing the computer-generated environment rather than the physical environment of the user. Necessary for experiencing presence are *involvement* and *immersion*, where *involvement* is defined as the “psychological state experienced as a consequence of focusing one’s energy and attention on a coherent set of stimuli or meaningfully related activities and events. Involvement depends on the degree of significance or meaning that the individual
attaches to the stimuli, activities, or events,” and immersion “is a psychological state characterized by perceiving oneself to be enveloped by, included in, and interacting with an environment that provides a continuous stream of stimuli and experiences. . . . A VE that effectively isolates users from their physical environment, thus depriving them of sensations provided by that environment, will increase the degree to which they feel immersed in the VE.”

In keeping with these definitions, and Witmer et al.’s claim that “the more control a person has over the task environment or in interacting with the VE, the greater the experience of presence” [193], users likely did not experience a sense of presence in the Sensorama because, while they may have been fully immersed—all of their senses except taste were stimulated—they probably were not involved due to their lack of control over what happened. Although the user grips the handlebars of the vehicle in Sensorama, the inability to physically move about also detracts from the feeling of presence, as “the more completely and coherently all the senses are stimulated, the greater should be the capability for experiencing presence. For example, adding normal movement, with kinesthetic [11] motion and proprioceptive feedback, should enhance presence.” [193]

In viewing a simulated environment on a standard computer monitor, users may perceive that they are on the outside looking in, which may detract from their sense of immersion, regardless of their level of involvement. [193] Therefore, inhabitance in virtual environments has originally been achieved through the use of a head-mounted display (HMD), a device that displays the virtual environment while simultaneously blocking all external visual stimuli. Since the user’s hands are visually absent, they must be explicitly rendered when necessary. Additional drawbacks of HMDs can include lags in image update time in response to rapid movement, such as when the user turns her heard quickly. Furthermore, lack of peripheral vision can cause dizziness after prolonged use. Another problem is that HMD interfaces are unable to reasonably accommodate more than one person, despite the assertion by Cruz-

Neira et al. that “For virtual reality to become an effective and complete visualization tool, it must permit more than one user in the same environment.”[33]

Another way of being immersed inside a synthetic environment is on the body scale, rather than the eye scale; the user’s whole body is physically inside a room-size display. One such system is the CAVE	extsuperscript{TM} [33], a cube measuring about 10 feet on a side, whose walls and floor are display screens. Participants may wear special glasses to view the stereoscopic images, if provided. A location sensor is used to enable the viewer’s perspective to be displayed on these screens. So, although the size and structure of the CAVE	extsuperscript{TM} enables it to accommodate several people at once, it is typically controlled by just the one tracked user, making all the others mere observers. Apart from allowing the inclusion of additional users in a minor role, CAVE	extsuperscript{TM}-like systems allow participants to see and use their own hands and other objects in the space.

Whatever benefits VR has for able-bodied users, they offer so much more for children with disabilities, as these systems can enable them to experience things the rest of us take for granted. “People with limited mobility can engage in activities in virtual space which they would not normally be able to participate in,”[119] in much the same way that Sensorama gave Rheingold a sense of riding a dune buggy.

The downside of the CAVE	extsuperscript{TM} is that the special-purpose hardware is costly, an inefficient use of space, and not easily transportable.

Now that we have introduced what virtual environments are, we will talk about the following relevant topics: virtual input devices, virtual output devices, and VE as an educational aid. We will then look more closely at a selection of VE/VR applications.

### 3.4.1 Virtual Input Devices

GUI-oriented input devices are obviously useless in any sort of virtual environment, as users are not typically seated at a table on which they can rest a keyboard or mouse; new input devices are thus crucial to the success of a VE. The most familiar of VR input devices is a
wand, a tracked joystick with buttons used to navigate and manipulate objects within the VE.[136]

The Cubic Mouse[47] is an input device similar to the mouse, but works in 3D. Unlike a regular mouse that is difficult to navigate within a 3D application, the cubic mouse can move independently along all three axes by the user’s pushing and pulling a set of three rods. Similar to it is a tool designed to help neurosurgeons view images of a model brain.[61] Rather than specifying x, y, z, yaw, pitch, and roll, which can be confusing, this system works within an augmented reality environment and allows the doctor to hold a doll’s head in one hand and a cutting-plane tool in the other. Rotating the physical head rotates the image on the screen, and by pointing the cutting plane-tool at the head, the system computes and shows the 2D image on that plane. Quite an intuitive tool, even non-neurosurgeons were able to use it effectively after about a minute of playing with the props.

Voodoo Dolls[123] is a technique for manipulating objects in a virtual environment using both hands. The user wears a pinch glove on each hand; these can sense when a user is “pinching” their fingers, and results in an effect on the object that it is determined they are holding. Being able to use both hands is key and “can be an advantage when operating controls in a VE, when the nondominant hand grips an interactive object (a segment of the virtual world) while the other hand carries out manipulations.”[167] Other gesture recognition approaches for use in VR are detailed in §3.5.3.

Surface Drawing[153] tracks a person’s gestures as a way of drawing 3D objects. The user must don special gear, including goggles and a glove. Since it is impossible to mark 3D space in the physical world, surface drawing is done in a semi-immersive virtual environment using the Responsive Workbench, which “present[s] the illusion that objects are floating in space.” This allows the user to see the drawing as well as objects in the physical world. The thumb defines the beginning and end of a stroke, by acting as a toggle switch; the stroke begins when the thumb closes up against the index finger, and ends when opened. Physical tools are available to make modifications to the drawing. Kitchen tongs are used to move, rotate,
and scale; a magnet tool deforms the shape (typically used to smooth out rough spots); and an eraser tool erases mistakes or cuts out small portions from the drawing.

Although Voodoo Dolls and *Surface Drawing* use quite a bit of technology, in essence, they both use the immersed person’s body as the input device. Peeling away these layers of technology allows the unencumbered body to serve as the primary input device, as we will see in Chapter 5.

### 3.4.2 Virtual Output Devices

In typical WIMP GUIs, with the exception of the rare force-feedback joystick and the like, characteristically the sight and sound senses are the only ones stimulated. While touch has mostly been neglected, “haptic feedback is generally considered to improve the quality of interaction in virtual environments” \[167\] since it adds to the sense of immersion and thus the feeling of presence. Youngblut \[197\] remarks that haptic feedback has a positive impact on education as well: “With respect to the use of multisensory cues, sound and haptic cues seemed to engage learners and direct their attention to important behaviors and relationships more than visual cues alone.”

In VEs that use HMDs, users hold real objects and “more or less believe that the object they are touching is a virtual one.” \[140\] For example, if the environment is a batting practice simulation, the user may hold something that feels like a bat in the hands, and the belief that it is indeed a bat is supported by the visual representation of the bat in the display. Such an approach is not normally used within a CAVE™ environment—being able to see their hands, it is harder to fool the user into believing they are holding something when they are indeed holding something else.

Special gloves which use pneumatics to provide feedback to the hand and fingers have been produced \[164\], but they do not appear to be commonly used. The creators of NewtonWorld \[35\] have added the use of a haptic vest, and regarding it, Youngblut concludes that, “Students appeared more engaged in activities when more multisensory cues were
For less immersive systems, PHANToM provides force feedback to a single point, letting the user feel if they are touching a virtual object: “If the position coincides with the position of a virtual object, the user feels a resisting force that pushes the tip of the pen back to the surface of the virtual object. Thus, by moving the pen, the user can trace the outline of virtual objects and feel them haptically.”

3.4.3 VR as an Educational Aid

Aside from VR’s value as an entertainment system, virtual reality holds enormous worth for education. VR can facilitate constructivist learning activities, which increase motivation and better resist mind-wandering, as children are immersed not only in the environment, but also in the activity. “In studies comparing immersive VR to two-dimensional desktop or even video instruction, the immersive users enjoyed their experience the most and reported the most desire to continue learning about the subject.”

Witmer et al. maintain that “many of the factors that appear to affect presence are known to enhance learning and performance . . . Factors believed to increase immersion, such as minimizing outside distractions and increasing active participation through perceived control over events in the environment, may also enhance learning and performance.”

In comparison with a similar paper-based curriculum that included lab time, students in an immersive environment learned as much or more.

Experiments on the multi-user virtual environment River City, a tool for teaching science concepts, found that typically weak students did as well as their academically stronger peers in the immersive environment; they were able to shed their real-world identities, and step into a successful scientist persona.

Another aspect of VR that makes it so useful as an educational tool is its ability to bypass time and size constraints. As with other computer simulations, we don’t have to wait for things to happen, we can see the results right away, and size can be adjusted to desired
dimensions: “While in a real garden children can learn how to plant, in the virtual garden they can learn how to think about plants, scale and position parts of an ecosystem, take on different roles, become small to observe the roots, talk and interact with children at distant locations, or factor time to directly observe the effects of their changes.”[136]

VR, as well as many non-VR applications, enables the learning experience to be finely tuned to the child’s needs and interests and provides alternative forms of learning which may support different types of learners, such as visually oriented learners.[197] In VR, however, the user is “an integral part of the stimulus flow, [so, if] meaningfulness and active control over a user’s experiences aids learning, then immersive environments likely are better training tools than standard computer-based training environments.”[193]

Roussou reminds us that it has not been shown “that VR can be used in a real-world educational context with non-expert learners on a long-term basis;”[138] teachers, as educators and facilitators, are still an important part of the education equation.

### 3.4.4 A Selection of VE Applications

Non-educational VE applications include Virtual Orchestra Performance[169], which has users conducting an orchestra in a VE, and Touching Space[180], a large-scale performance environment in which dancers’ movements create augmented visual effects and are viewed by the audience with polarized glasses. KidsRoom[14] is an environment for kids whose physical space is augmented with “graphics, video, sound effects, light, music, and narration.” In this environment, children interact with physical objects in the space (they talk to the furniture and move a bed around), virtual creatures that are projected onto the walls (they dance with monsters), and each other (they move through a forest and row a boat). Without the required collaboration, the narrative stalls momentarily while it tries to persuade everyone to work together.

Nautilus[166] is another collaborative virtual reality game. In this game, players control
the movement of a diving bell\footnote{A type of diving apparatus which, incidentally, did not allow divers to control its movement.} in their mission to rescue a dolphin trapped beneath a shipwreck. The players’ movements are tracked by sensors in the floor, which dispenses with the need to wear special gear, and compels collaboration, as “direction and speed are controlled by the group’s centre of mass.”

Educational VR applications are discussed in §5.1

3.5 Tracking

Enabling a computer to track objects and make sense of users’ positional information as well as their actions, and give appropriate feedback is one step along the path toward conquering the great divide between bits and atoms that Ishii and Ullmer\footnote{Ishii and Ullmer} refer to. Numerous schemes and technologies have been proposed by researchers to accomplish this, which is discussed here.

3.5.1 Tracking Objects

There are many different aspects of an object that might require tracking, such as position, orientation, and motion, but unless there is just a single object, we need to identify the trackable objects. Many of the early systems used direct, wired connections to identify its components. The pieces had to be specifically designed and built, to work with the system, with internal processors to allow adjacent pieces to communicate with each other and send information back to the host computer. Computational Building Blocks\cite{2}, Triangles\cite{54}, and ActiveCube\cite{91} all work in this way, incorporating processors into the custom-built pieces to recognize neighboring pieces.

While the aforementioned implementations are all wired, with a base onto which other pieces must be attached in order to receive power and be counted as part of the configuration, Resnick et al.\cite{131} states that wires limit “not only what children can build but also how they
think about their constructions.” Newer systems incorporating microcontrollers communicate wirelessly, via radio or infrared communication. Wireless communication among components and a host computer takes care of Resnick’s concern and removes other restrictions inherently imposed by a wired configuration, such as entanglement, movement restriction and the single route/bottleneck to the host.

However, wireless communication has its own considerations. While communication across wires can be shielded from interference, communication across airwaves can not. Of the two types of wireless communication, infrared (IR) and radio, infrared still restricts movement somewhat, as it has a line-of-sight requirement. This requirement, however, makes it less susceptible to interference: potential interference by IR emitters in the vicinity is not an issue unless directly in the line-of-sight. Without the same requirement, radio communication is an option for more applications as it does not limit movement, yet it can be interfered with from nearby devices, regardless of position or placement. To combat interference, additional error-checking must be included to filter out communication among other devices.\[67\]

Additional concerns to be taken into account when using wireless communication with microcontrollers include the electromagnetic radiation given off by the devices as well as the need for an onboard power source (wired technologies can draw power from the host), which adds weight and may complicate debugging during design.

While microcontrollers allow for the integration of sensors\[13\] for acquiring additional information, barcodes, computer vision, and radio frequency IDs can be used to wirelessly identify objects that do not require this additional capability. Some systems, such as an add-on to the KidPad\[37\] collaborative drawing tool, use barcodes\[163\] for identifying objects. Barcodes are easily recognized by a barcode reader, and can be generated on a computer and printed with a standard printer. However, barcodes are of limited use, as they are only good for identification.

\[13\]Such as gyroscopes, compasses, inclinometers, and accelerometers.
<table>
<thead>
<tr>
<th>Communication technique</th>
<th>Advantages</th>
<th>Disadvantages</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wired</td>
<td>• no external power source required</td>
<td>• limited to a single construction</td>
<td>• Triangles</td>
</tr>
<tr>
<td></td>
<td>• minimal interference</td>
<td>• movement restricted</td>
<td>• ActiveCube</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• PlayPals</td>
</tr>
<tr>
<td>Wireless (infrared)</td>
<td>• mid-level movement restriction (must remain within view of IR sensor)</td>
<td>• line-of-sight required</td>
<td>• Navigational Blocks</td>
</tr>
<tr>
<td></td>
<td>• interference only from IR emitters within view</td>
<td>• generates radiation</td>
<td>• TV remotes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• additional coding necessary to ensure reliability</td>
<td>Wii™ remote</td>
</tr>
<tr>
<td>Wireless (radio)</td>
<td>• no line-of-sight required</td>
<td>• generates radiation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• least restricted movement</td>
<td>• interference from all directions</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.1: Comparison of communication techniques used in TUIs with internal processors.

Computer vision, using cameras connected to the computer to view the environment, can give the computer a larger and more complete picture. While computer vision can work without identifying tags, changes in lighting and other factors can distort the image, making it difficult to identify arbitrary objects. Fiducials\textsuperscript{14} can be used on objects to more easily identify them through computer vision. Like barcodes, fiducials can easily be generated, distributed, and printed, but unlike barcodes which only identify an object, computer vision sees the whole picture and can therefore identify object rotation as well. Fiducials can also be encoded with data about the object.\textsuperscript{142}

Although the earliest systems, like DigitalDesk\textsuperscript{190} and I/O Bulb\textsuperscript{174}, utilized overhead cameras for tracking objects on a tabletop, computer vision is susceptible to occlusion. People bending over the table to pick up or move an object would block the camera from viewing the objects, which interfered with the tracking. Computer vision is also subject to noise on the image, which can result from motion, changes in lighting conditions, and dust, among other things. To combat these two problems, the cameras were moved beneath the table, and a semitransparent surface was used as the tabletop\textsuperscript{15}. Putting the cameras out of

\textsuperscript{14}Anything taken as a standard of reference, such as an identifying tag.

\textsuperscript{15}This can also be used as a projection surface, by placing the projector beneath the table as well, which then also removes the projection-occlusion issue.
the way effectively reduces occlusion to zero as long as the tabletop is kept cleared, while the
diffuse surface blocks out all objects not sitting directly on the surface\textsuperscript{16}, while maintaining
more even lighting. With these two modifications in place, however, fiducial markers are a
must, as even shapes on the table are slightly blurred. These fiducial tags, while typically on
the underside of objects being tracked, still visually modify the objects, which may be seen
as a shortcoming in some applications, such as when an object may be flipped on its side.

Systems that use computer vision include Underkoffler and Ishii’s urban planner Urp\textsuperscript{17} and
Baudisch et al.’s Lumino blocks\textsuperscript{12}. These systems are covered in more depth in the
next chapter. Kaltenbrunner et al.’s reacTIVision\textsuperscript{8} uses computer vision to track objects,
including fingertips, that have been tagged with fiducial markers, as they are moved about
on a tabletop.

Although computer vision works fine for two-dimensional applications (where the camera
needs to see just one side of an object), especially when placed in a position where occlusion
is minimized, it is not usually sufficient for 3D applications, which are subject to occlusion
by definition.\textsuperscript{17} Radio frequency ID (RFID) tags is another useful tool for tracking tangible
objects, especially when line-of-sight can not be guaranteed.

RFID tags contain an integrated circuit for storing data as well as a radio antenna, and
can be passive, semi-passive, or active.\textsuperscript{18} An RFID reader is used to detect the presence
of the RFID tags. The frequency of the RFID combined with its passivity dictates how
close it has to be to the reader in order to be read. RFID tags are quite small, and can be
embedded in an object without modifying its appearance. However, the hardware can not
be generated by a standard printer or distributed by e-mail, and although position can be
computed, rotation cannot.

Because RFID does not inherently provide location information, it is best used when

\textsuperscript{16}This minimizes variation in object distance from camera for less error-prone tracking.

\textsuperscript{17}Depth cameras, which can provide some 3D information, are by no means occlusion-proof—they are not
capable of seeing inside or beyond an object.

\textsuperscript{18}Passive RFIDs do not use batteries, semi-passive RFIDs have an internal battery for powering their
circuits yet use the RFID reader’s power to broadcast their signals, and active RFIDs continuously broadcast
their signal.
positioning of objects is constrained by the system: DataTiles\textsuperscript{[130]} and Senseboard\textsuperscript{[75]} both use RFIDs on tangible objects constrained to placement in a grid, and each grid position reads the RFID of the object placed there. Although RFIDs can be subject to interference from nearby metal, this can be put to good use as in Ishii et al.’s musicBottles\textsuperscript{[71]} which use the interference to modify the signal when a bottle is opened or closed.

Privacy is a concern with RFID tags, since they may be detected from a distance, without the knowledge or consent of the person whose possession it is in, and may not even be aware that they are in possession of such a trackable item. This genuine concern must be addressed by those researchers using RFID.

To combat the bulk of the shortcomings in both computer vision and RFIDs, Olwal and Wilson combined both in SurfaceFusion\textsuperscript{[114]}, using RFID to detect the tangible objects, and computer vision for obtaining position, shape, and movement. This research is promising but has concerns unique to a combination approach: maintaining object registration across sensor systems and incorporating both sets of sensors so that they do not interfere with each other.

Although wireless identification technologies allows the freedom to tag any standard object for use in a TUI, the pre-tagging requirement still limits the set of usable objects to those that have been tagged.

Aside from merely knowing that an object is present, there are numerous other properties which might be necessary for the computer to track, such as position, orientation, and motion. Wired six degree of freedom (6DOF) sensors were used in early systems like Ullmer and Ishii’s metaDESK\textsuperscript{[173]} and Fitzmaurice et al.’s Bricks\textsuperscript{[45]} to track the 3D position and orientation of objects (active lens and bricks, respectively). Although 6DOF sensors give absolute position information, and are useful for tracking people (see next section), for objects, we often need to know placement relative to one another—if they are touching, and how. This information is typically obtained through contact leads on TUIs with embedded microcontrollers or computer vision techniques.
<table>
<thead>
<tr>
<th>Identification technology</th>
<th>Advantages</th>
<th>Disadvantages</th>
<th>Application</th>
</tr>
</thead>
</table>
| Microcontroller            | • Active components may do some processing  
• Good choice for 3D constructions | • Requires large degree of customization  
• Individual components require power source if not wired | • Triangles[54]  
• ActiveCube[91]  
• Computational Building Blocks[2] |
| Barcode                    | • Easily identifies the piece  
• Easily generated and distributed | • Can not detect rotation or position  
• Visually modifies the object | • KidPad add-on[163]  
• WebStickers[96] |
| Computer vision             | • Can detect rotation of 2D objects  
• Fiducials can be generated easily and distributed  
• Can track untagged objects | • Subject to occlusion  
• Subject to noise on image  
• Fiducials visually modify object  
• Direct line of sight necessary | • Urp[175]  
• reacTIVision[81]  
• Lumino[12]  
• SurfaceFusion[114] |
| RFIDs                      | • Very small  
• Direct line of sight unnecessary  
• Does not change appearance of object  
• Fast, reliable | • No rotation information  
• Can not be generated or distributed electronically  
• Privacy concerns  
• Interference | • Senseboard[75]  
• DataTiles[130]  
• musicBottles[71]  
• SurfaceFusion[114] |

Table 3.2: Comparison of common identification technologies used in TUIs.

Computer vision can locate objects by comparing an empty scene with the present scene or by looking for certain colors. Although tracking with computer vision is typically limited to two dimensions, Baudisch et al.[12] incorporate fiber optics ingeniously into a computer vision system to track multiple objects being manipulated and stacked on a tabletop. Allowing light to pass through permits the fiducial markers of objects above to be seen, and join with the marker of the current object to form a unique tag which defines how pieces are stacked. Drawbacks include constraint of rotation and declining clarity as the number of objects stacked increases.
3.5.2 Tracking People

Many of the same methods for tracking objects can be used for tracking people, but since tracking people is often a precursor to gaining additional information about the person (such as their pose or gestures), we take another look at the relevant technologies, with people in mind. Of the technologies listed in Table 3.2 on the preceding page, barcodes can detect neither absolute nor relative position, and microcontrollers are typically used for identification of other objects that they touch—they can detect relative position (and sometimes orientation) of adjacent objects.

One of the most well-known tracking systems is the Global Positioning System (GPS). To locate position, GPS receivers acquire radio signals from a minimum of four satellites, to determine longitude, latitude and altitude.\(^\text{19}\) Accuracy of GPS can be affected by a number of factors including atmospheric conditions, and is generally not fast enough for most applications. Moreover, it may be difficult to acquire enough satellites indoors or in areas where there are many tall buildings.

Menache\(^\text{101}\) describes a Local Positioning System (LPS) prototype which works in a similar manner, albeit having the tracked person wear a device which transmits radio signals rather than receives them. An array of antennas (in place of the satellites) surround the area—up to \(25,000\text{m}^3\) in size—to receive the transmissions. The relatively large size of the tracked area makes this a feasible technology for tracking people playing sports. LPS signals, however, can not travel through metal surfaces, thereby precluding it from some applications. Furthermore, as with all devices that depend on radio communication, there is the opportunity for coincidental or malicious interference.

The 6DOF sensors mentioned in the previous section take a similar approach. Users wear transmitters which generate low frequency electromagnetic fields. These electromagnetic fields are detected by nearby receivers and decoded into position \((x, y, z)\) and orientation (yaw, pitch, roll).\(^\text{101}\) Analyzing position and orientation as they changes over time

---

\(^{19}\)See \cite{55} for an easy-to-understand explanation of how this works.
should make it possible to recognize some gestures. Attaching more than one sensor to a person would allow more gestures to be recognized but, if wired, the wires would constrain movement, and if wireless, the transmitters would need an on-board source of power, significantly heavier than a passive marker. While the data they provide is comprehensive, the trackers are usually wired and sensitive to nearby metal. Also, the trackable area is smaller than that of other systems and there is concern over being exposed to electromagnetic radiation.

Optical systems are often used to track people, as they “offer the performer the most freedom of movement since they do not require any cabling.”[172] In addition, Maes et al.[97] found that users may “enjoy greater behavioral and expressive freedom,” and observed some users of their ALIVE system feeling free to do cartwheels, jumping jacks, and the like. Moreover, users were able to concentrate more on interacting with the environment itself, rather than focusing much of their attention on the complex and unfamiliar interaction equipment that some environments require. On the downside, vision systems are “limited by the resolution of the cameras and the sophistication of their tracking software.”[172]

Early optical systems used standard cameras to track users, and because of the minimal equipment demands, researchers continue to investigate their usage. You et al.[196], for example, use computer vision to track people in a living room to determine how the screen real-estate of a large-format TV is divided into individual areas, such that each person’s allocated area is positioned relative to where they are in the room, i.e., if person A is sitting between person B to the right and person C to the left, the screen will be divided into three areas, C, A, and B, from left to right. If people switch places, so too does their space on the TV.

More often, newer vision systems use reflective markers placed at strategic locations on the person being tracked. These markers correspond to key points on a computer model of the person. Then, using a minimum of three cameras, the 3D position of each one of these

---

20Since the ALIVE system is primarily concerned with tracking people’s gestures, it is more fully explored in §3.5.3.
markers can be calculated via triangulation. To counteract occlusion, most systems use more than the minimum number of cameras and/or place additional markers on the performers.\footnote{Neither of these are without shortcomings, as “adding more cameras makes tracking each marker more complex, resulting in increased CPU time. Increasing the number of markers can result in exponentially increasing the “confusion factor”, i.e. keeping track of which marker is which.”\cite{172}} Tracking multiple performers at once requires putting markers on all of them, and mapping these to separate models.

The Vicon motion capture system takes this approach, projecting infrared light onto the environment, which is reflected by the markers.\cite{179} The cameras filter out all-but-infrared light so that the system sees only the markers; background objects are therefore filtered out. Since infrared light is used to help illuminate the reflective markers, they can be used outdoors (the sun’s infrared light provides additional illumination), and indeed, Vicon makes a system optimized for outdoor use.\cite{178} However, these systems work best with other light sources minimized, and black attire is recommended.

Microsoft’s Kinect\textsuperscript{TM} sensor for the Xbox\textsuperscript{R} gaming console is a motion sensing input device which uses computer vision, albeit with “an infrared\textsuperscript{22} projector combined with a CMOS sensor\textsuperscript{23},” to measure depth, according to the Kinect Fact Sheet\cite{88}. \footnote{The type of infrared used is believed safe for humans, but since scientific tests on prolonged usage has apparently not been done, there is still concern about exposure.} \footnote{A CMOS sensor is a kind of image sensor.} \footnote{Therefore it is technically not true tracking, but since it locates the person and identifies body parts, it is included here.} Similar to echolocation, where sound is bounced off objects and time to echo is measured, infrared depth cameras like the Kinect project infrared light over the environment and build a depth map from the length of time it takes the light to return, with additional information from the deformation of the projected light.\cite{25} The computed depth map contains position data for all people and objects in the environment.

While most other systems require a calibration pose to initially locate body parts, and track in part by assuming body parts could not have moved too far from one frame to the next, the Kinect works on a frame-by-frame basis\footnote{Therefore it is technically not true tracking, but since it locates the person and identifies body parts, it is included here.} without the use of any temporal information. This obviates the need for a calibration pose. Other advantages of this approach
are its resistance to loss of tracking due to occlusion and its ability to handle multiple users well when they can all be clearly seen by the sensor, such as in a typical gaming situation. By contrast, drawbacks of the Kinect include: in collaborative applications that more closely resemble the real world, with teammates moving about, often blocking other team members, participants may become occluded from the sensor. According to the Kinect Manual, lone players must maintain a minimum distance of six feet from the sensor, while two people should keep at least eight feet back.[87] Distance for larger groups is not discussed. While infrared depth perception should not be affected by varying light levels, the manual suggests “turning on lights to brighten the play space,” if the sensor does not see the player. Shining lights (particularly sunlight) onto the sensor, on the other hand, can be a problem as well, since the infrared rays emitted by the external light sources will be jumbled with the reflected infrared rays the sensor is expecting.

Table 3.3 on the next page compares the various technologies available for tracking people’s position and/or orientation.

Some researchers have come up with low-tech ways to track users or handle tracking confusion. Although KidsRoom[11] restricts the number of people-to-track to four and objects-to-track to one to attempt to avoid confusion, it may still happen. The researchers mention an interesting way to cope with mix-ups of multiple users in VEs: “unlike conventional surveillance tasks, an immersive environment that must keep track of identity can manipulate people in its environment so that when it becomes uncertain of identity it can ‘bootstrap’ itself automatically. For instance, a system controlling an environment with a telephone might physically call the room, ask to speak to a particular person and, when that person comes to the phone, reinitialize tracking.”

Tracking of body parts, rather than the full body may help those with special needs. Researchers have investigated assistive tracking for people with multiple sclerosis and muscular dystrophy, and have had success with electromagnetic trackers placed on the hand. The trackers, combined with special predictive software, allowed those with minimal
<table>
<thead>
<tr>
<th>Approach</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
</table>
| radio frequency positioning systems (e.g., GPS) | - can work outdoors  
- some may work indoors  
- works over relatively large area  
- can track many people | - signal occluded by large or metal surfaces  
- possible RF interference  
- can not track gestures or orientation |
| electromagnetic trackers | - not vision-based, so not subject to occlusion  
- can track multiple people  
- tracks both position and orientation | - typically wired  
- limited ability to track gestures  
- small capture area  
- sensitive to metal  
- concern over exposure to electromagnetic radiation |
| standard computer vision    | - wireless  
- able to track movement/gesture as well as position  
- subject to occlusion | - changes in conditions can affect tracking  
- subject to occlusion  
- restricted to environment  
- may require an initial pose  
- works best with individuals |
| reflective computer vision (e.g., Vicon) | - wireless  
- only sees trackable reflectors  
- effect of noise is lessened  
- markers may be reconfigured easily  
- able to track gestures too  
- can be used outdoors | - require users to wear reflectors  
- reflectors can still be occluded particularly with few cameras  
- restricted to environment  
- reflectors may be confused with one another  
- hardware can be very expensive  
- portability is limited  
- environment must be controlled (black background, black clothes)  
- possible concern over infrared exposure |
| depth cameras (e.g., Kinect) | - wireless  
- affect of varying light levels is significantly reduced  
- no fiducials required  
- calibration pose not required  
- able to track gestures too | - significant minimum distance to sensor is required  
- concern over infrared exposure  
- may not work well for collaborative setups  
- can not be used outdoors (interference from sunlight)  
- restricted to the environment |

Table 3.3: Comparison of technologies for tracking people’s locations and/or orientation
arm or finger control to interact with a computer.\cite{197}

We have discussed tracking people within virtual environments, because that is what we are primarily concerned with, but tracking can be done in the general world as well, as in the Whereabouts Clock\cite{155}, which uses an application running on participants’ cell phones to keep track of their location.

Many systems which track people within VEs are also concerned with understanding their gestures, and these are discussed in the next section.

### 3.5.3 Gesture Recognition

According to McNeill\cite{100}, gestures add another dimension to speech, and they exist on a continuum from gesticulation to structured sign language for the deaf, with gesticulation being the most unformed and non-reproducible. Although gesticulations may be impossible to decipher, at times even to humans, other gestures are structured enough to be understood by a computer.

As we mentioned earlier, in moving away from WIMP GUIs, researchers are introducing new modes of interaction, including gesture-based interfaces. An early form of gesture recognition is an on-the-spot handwriting recognition program like Palm\textsuperscript{TM}'s Graffiti\textsuperscript{R}\textsuperscript{25}, which works in two dimensions, on a touch surface. Multi-touch displays take things a bit further by recognizing 2D gestures with multiple simultaneous points of input. While this is convenient for actions that are typically done on a surface, most actions are not, but rather occur in 3D space.

There are two main approaches to gesture recognition, either requiring the user to hold and/or manipulate a device, or using computer vision to see and extrapolate the user’s gesture. Some researchers have tried using a combination of these approaches.

Pinch gloves, as used in Voodoo Dolls\cite{123}, are a set of specialized gloves worn by the users, with sensors on the thumb and pointer finger which enable a “pinching” gesture.

\textsuperscript{25}Unlike an OCR program which works on pre-written text, Graffiti tracks the movement of the stylus on the touchpad.
(only) to be sensed. Typically a motion tracker is attached as well, to give positional information, so that the context of the pinch gesture could be understood. The pinch gloves themselves are not very comfortable, nor is the gesture itself a very natural pinch, which may interfere with the usability of the gloves. Users of different hand size and finger length may require differently sized gloves. The fine motor control required for the pinching motion may also be too difficult, or even impossible, for people with handicaps or little kids. On the plus side, multiple users can be fitted with pinch gloves and work collaboratively within an environment.

Handheld devices capable of recognizing the user’s gestures are not worn, but rather held, and have found their way into many households already. Nintendo® has introduced the Wii™ game platform with a handheld remote control that uses accelerometers to enable games to recognize the gestures of one or more players. In the game of Wii™ Sport, for example, gamers may swing a golf club, tennis racket, or baseball bat, or throw a bowling ball. The accuracy of the swing or throw affects game play. Johnson et al. [77], too, recognizes complex gestures performed on a child’s plush toy. Used to control an on-screen character, the toy can be moved in various ways, which translate to run, fly, kick, etc. Both of these handheld devices, however, only recognize active gestures; they have no way to recognize gestures like pointing.

Pinch gloves and handheld devices do not suffer from the drawbacks of vision-based techniques, thereby not being subject to occlusion or changes in environmental conditions.

*Flashlight Jigsaw* [23] combines a handheld device with computer vision. *Flashlight Jigsaw* is a jigsaw-type game played on a large-screen projected display by up to three players holding controllers. While these controllers are spatially tracked by the Vicon motion capture system, buttons on the device make additional functions available to the user.

While handheld devices allow for the inclusion of additional controls, such as buttons, and the available device-based techniques are capable of recognizing quite a few gestures, we are primarily interested in the simple, universally understood gesture of pointing. According
<table>
<thead>
<tr>
<th>Approach</th>
<th>Benefits</th>
<th>Drawbacks</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pinch gloves</td>
<td>• not vision-based, so not subject to occlusion or changes in environment</td>
<td>• tethered 6DOF tracker</td>
<td>Voodoo Dolls[123]</td>
</tr>
<tr>
<td></td>
<td>• can track multiple users</td>
<td>• unnatural feel</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• unnatural pinch gesture</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• gloves must be available in a variety of sizes</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• may not be adaptable for handicapped users</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• restricted set of gestures</td>
<td></td>
</tr>
<tr>
<td>Handheld Device</td>
<td>• not vision-based, so not subject to occlusion or changes in environment</td>
<td>• requires user to hold device</td>
<td>Sympathetic Interface[77],</td>
</tr>
<tr>
<td></td>
<td>• device may operate much like a joystick</td>
<td>• can only recognize active gestures (not pointing)</td>
<td>Nintendo Wii™</td>
</tr>
<tr>
<td>Handheld + Vision</td>
<td>• handheld device allows users to easily make a selection by pressing a button</td>
<td>• requires user to hold device</td>
<td>Flashlight Jigsaw[23]</td>
</tr>
<tr>
<td></td>
<td>• built-in sensors allow easy rotation of puzzle pieces (simply rotate the controller)</td>
<td>• pointing direction is calculated from the angle of the device in the user’s hand rather than user’s true pointing direction</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• requires vision for handheld’s position</td>
<td></td>
</tr>
<tr>
<td>Vision</td>
<td>See Table 3.5 for an in-depth comparison of vision-based gesture recognition approaches</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3.4: Comparison of device-based gesture recognition approaches

to McNeill[100], pointing is a deictic[26] gesture, which has “the obvious function of indicating objects and events in the concrete world, but it also plays a part even where there is nothing objectively present to point at.” Störring et al.[165] further state that such a gesture has two main uses: “to indicate a direction or to pinpoint a certain object.” Both of these uses are of interest in virtual environments: pointing a direction to turn in a virtual tour, or indicating or selecting a virtual object. For this, a handheld controller may not be the best, or even a good, choice. Understanding human’s true, device-less, three-dimensional gestures is more of a challenge.

---

26 According to Merriam-Webster (www.m-w.com), deictic means “showing or pointing out directly” like the words this, that, and those.
Some issues facing researchers when tracking gestures with computer vision, are the low ambient lighting necessary for clear viewing of the projected display, and correct camera positioning to avoid occlusion. Often, fiducials (identifying tags) or markers (typically reflective balls marking a point) are used. Fiducials are recognizable on sight, while markers must be understood in the context of any other markers, but the reflectiveness of the markers can more easily be seen in low lighting. Calibration is often required whether or not fiducials or markers are used.

SixthSense\cite{104,105} is a relatively simple single-camera wearable computer-vision gesture tracking system that uses natural hand gestures to interact with augmented artifacts. Fiducials are used to mark the user’s fingertips and other items in the environment, which are then “interpreted into gestures that act as interaction instructions for the projected application interfaces.” Multi-user and multi-touch interaction are constrained only by the number of unique fiducials.

Maes et al.’s ALIVE system\cite{97} uses just one video camera to obtain video where pattern recognition combined with color space classification is applied to determine the position of the head, hand, and other relevant features, which must be against a fixed background. They composite this image into a 3D virtual world that “is projected onto a large screen that faces the user and acts as a type of ‘magic mirror’: the user sees herself surrounded by objects and agents.”\cite{27} Within that environment, the user’s location and gestures affect the behavior of these agents, which, in turn, provide visual and auditory feedback to the user. Gestures, including pointing, are recognized in two dimensions, and it appears that for clarity, gestures are expected to be made with the hands in the same $z$ plane as the rest of the body.\cite{28}

Computer vision may also be used to track pointing in three dimensions, using more than one camera, but often requires various constraints, such as the use of fiducial markers, restricting pointing direction to a single plane, restricting the number of users, or restricting

\footnote{In artificial intelligence, an \textit{agent} is an autonomous entity within an environment, which observes and takes action toward achieving a goal.}

\footnote{For example, pointing forward to the left would be seen as pointing to the left.}
clothing color. Leubner et al.\cite{Leubner2006}, while not using fiducials, works for a single user, mandates the user point only at the display, must be a specific distance away, and requires the background to be static.

Störring et al.\cite{Stoerringer2009}, also tracking just a single user, utilize the position of a wired sensor attached to stereo glasses\footnote{Worn to view the 3D affect in the VE.} as the position of the head, and track the skin-colored blob that is not at the position of the head, as the hand. As long as the user is not wearing red clothing, the head and hand can be found. For Caucasian users, this approach works when combined with the position of the head tracker, if the displayed images are not too dark and no one color is predominant. To help the user, “[he] can see his pointing direction indicated by a 3D line starting at his hand and pointing in the direction the system ‘thinks’ he is pointing. Thus, the user can adjust the pointing direction on the fly.”\cite{Stoerringer2009}

Kehl et al.\cite{Kehl2010} labels the extreme edges of a single user’s silhouette as their head and pointing fingertip to compute pointing direction, which they define as “the line of sight connecting the eyes and the fingertip of the pointing arm.” Drawbacks of this technique involve the extension of an elbow or leg extends which is therefore mistaken for the head or hand, and lack of support for pointing at the floor. Carbini et al.’s SHIVA\cite{Carbini2009} allows dual-user gestural interaction in front of a large display, but they too, restrict pointing to the specific plane of the display.

The biggest problem inherent in a gesture-based interface is fatigue. Some systems, such as the one described by Störring et al.\cite{Stoerringer2009}, require the position of the hand to remain constant for a period of time in order for the pointing gesture to be recognized. Not only can this lead to unwanted selections and slow down interaction, it can be extremely exhausting. Implementations that did not require the pose to be held were also deemed tiring if played long enough\cite{Cabral2009}. Cabral et al.\cite{Cabral2009}, recommend that “a gesture interface should be used as an alternative to existing interface techniques or complementing them.”

Microsoft’s Kinect allows users to interact with the system through gestures (and spoken
commands), no game controller necessary. A wide variety of gestures are recognized without the need for fiducials, but users must face the sensor. Using machine learning, Kinect is trained to recognize body parts, and thus gestures, from over a million images.

While Kinect can track articulated movement for multiple people simultaneously, and does not require holding a pose for any length of time, the single sensor is insufficient to recognize gestures of multiple participants in non-gaming setups, i.e., when players are not standing side-by-side facing the sensor. Another drawback of Kinect for collaboration: the number of actively gesturing players that can be tracked simultaneously is limited to two.

Using multiple Kinects has been proposed as a way to get around these limitations but, according to Microsoft, there would likely be interference between the Kinects’ infrared light sources, which reduces accuracy and precision. Microsoft also reminds us that, “Skeleton Tracking works best when the tracked user is facing the sensor,” yet users can only face toward one sensor at any single moment. On the programming end, “the Skeleton Tracking API does not synchronize skeleton IDs across multiple sensors.” Research in this area is ongoing.

See Table 3.5 on the following page for a comparison of the benefits and drawbacks of the various vision-based gesture recognition approaches we have discussed.

### 3.6 Summary

In this chapter we have discussed user interfaces, input devices, tangible user interfaces, and virtual environments. We have looked at a variety of examples and approaches of each of these, and explored their pros and cons. In addition, we have also explored the various methods of tracking the objects and people that the interface must be aware of. Finally, we have laid the groundwork for our research and now proceed to detail our advancements in the field. In the next chapter I discuss my work within the first line of research, tracking two-

---

30 Machine learning is a branch of Artificial Intelligence in which the computer learns, from data that has been fed into it, how to recognize complex patterns, so that it can make intelligent decisions for future data.
<table>
<thead>
<tr>
<th>Paper</th>
<th>Description</th>
<th>Benefits</th>
<th>Drawbacks</th>
</tr>
</thead>
</table>
| Maes et al.’s ALIVE[37]       | A single camera tracks a user interacting with an agent.                   | • no need for fiducials or markers  
• recognizes multiple gestures using pattern recognition                                      | • two-dimensional pointing  
• restricted to a single user  
• assumes a fixed background  
• user must face camera/screen |
| Störring et al.[165]           | Pointing gesture is estimated as the line beginning at a point between the eyes extending through the pointing hand. | • using computer vision and a magnetic tracker allows much error-checking                      | • requires the use of a magnetic tracker on user’s head  
• gesture not recognized until the position is held for a number of frames  
• tracks a single user |
| Kehl and Van Gool[85]          | Pointing direction defined as line-of-sight connecting the eyes and the pointing fingertip. | • no initial pose  
• no fiducials/markers needed  
• allows pointing in any direction except the floor                                           | • tracks a single user                                                                 |
| Carbini et al.’s SHIVA[24]     | A stereo camera tracks two users by skin color and movement tracking. Uses the head-to-hand line for determining pointing direction. | • no need for fiducials or markers  
• tracks multiple users  
• voice recognition and application context used to enhance gesture recognition                | • must face and point toward a specific plane  
• appears the users must take turns rather than work collaboratively  
• bimanual selection |
| Mistry and Maes’s SixthSense[104] | An augmented reality wearable interface which allows users to interact with the system using natural gestures. | • same POV as the user  
• supports multi-user and multi-touch                                                            | • requires the use of fiducials on the person’s fingertips as well as on trackable objects in the environment |
| Microsoft’s Kinect             | Uses structured light to compute depth map, and machine learning to recognize body parts and gestures. | • fiducials/markers unnecessary  
• not affect by light levels  
• multiple users supported, but with limit                                                     | • must face sensor  
• minimum distance to sensor is required  
• multiple users may occlude each other                                                            |

Table 3.5: Comparison of vision-based gesture recognition approaches
and three-dimensional tangible objects, and in the following chapter, the tracking research I’ve done within a virtual environment.
Chapter 4. Tangible Interfaces for Collaborative Learning Environments (TICLE)

As described in the Introduction, many educators advocate the use of tangible objects for learning: actively manipulating an object engages the brain in ways listening to a lecture does not. TUIs—tangible objects with a computer interface—are believed to possess additional educational advantages. Programming the computer to be aware of how students are manipulating a puzzle, and to give feedback, allows the computer to keep the children on track. Enabling a tangible user interface system to be useable by a group as well as individually, makes available the benefits of collaboration within the environment and expands the educational possibilities. As we discussed in §3.3 providing multiple access points supports collaboration by distributing control amongst participants and permitting shy people to be more actively involved.

Combining these computer-trackable manipulatives with intelligent tutoring in a collaborative setting is the focus of Scarlatos’s Tangible Interfaces for Collaborative Learning Environments (TICLE) research group[143], which included myself. We concentrated on ways computers can enhance learning by responding to students’ actions in a physical environment.

\[\text{1See chapter 2}\]
4.1 Two-Dimensional Tangible User Interfaces

To begin, we focused on tracking the usage of educational manipulatives with an inherently two-dimensional structure, such as standard puzzle pieces. Although many tangible objects are in fact three-dimensional, like Cuisenaire® rods or commonplace items like rocks, some can be looked at as two dimensional: the important aspect of the Cuisenaire® rods are their length relative to the other rods, and items like rocks can often supply relevant attributes—even about a third dimension—by encoding them in a 2D format. (See Scarlatos[142] for more on identifying tags.)

The benefits that educational TUIs can offer prompted the TICLE group to build an enhanced version of the tangram, an old Chinese puzzle. Two-dimensional in nature as well as appearance, the tangram is an excellent candidate for computer enhancement. The prototype system would watch children as they play with the tangram (see Figure 4.1) on a physical tabletop, and acts as a guide-on-the-side by offering encouragement and hints at appropriate times.[142] [146]

![Figure 4.1: The seven pieces of the tangram, an old Chinese puzzle.](image)

The seven pieces that comprise the tangram can be used to form many different shapes, with the most difficult one being a square, for which there is just one solution, as seen in Figure 4.2(a). The popularity of the tangram is due to its varied usages: helping the

---

2In fact, Cuisenaire® rods made for overhead projector use are 2D.

3The task requires that children focus on the form of the pieces and not their color.
very young learn the names of some standard polygons, allowing children to understand the concepts of area and congruence\footnote{Equality of size and shape.} without numbers or formulas, and developing a geometric intuition which should lay the foundation for grasping more difficult geometric concepts later on.\footnote{Equality of size and shape.}

4.1.1 Related Work

At the time of TICLE’s first published paper\cite{143}, in 1999, there were few related research projects that had been published. Underkoffler and Ishii’s \textit{Illuminating Light}\cite{174} is one
of the first research projects to track tangible objects and display related information. Intended to serve as a ‘simulated optics workbench’ for optical engineers to simulate working with expensive optics on a tabletop, *Illuminating Light* uses an overhead *I/O Bulb*—a combination camera-projector—in conjunction with vision analysis to recognize unique dot patterns on the components. While the physical components of the system were reminiscent of their working counterparts in appearance, the cameras simply tracked their location and orientation in 2D space.

*Urp*[^75], also by Underkoffler and Ishii, is an urban planning application that allows urban planners to see the effects of various arrangements under different circumstances. Placing architectural models on the work surface causes shadows to be cast that are “accurate for arbitrary times of day” as set by the user. *Urp* extends the concepts introduced with *Illuminating Light* by adding assorted tools which allow users to modify various attributes. Aside from controlling time of day, touching a model with a wand alternates between applying a glass and brick façade (a glass façade throws reflections while a brick structure does not), and setting a wind tool on the table generates a visual of pedestrian-level airflow that “takes into account the building structures present, around which airflow is now clearly being diverted.”

Both *Illuminating Light* and *Urp* consist of a number of items manipulated on a tabletop, thereby opening up the interaction to multiple concurrent users and allowing them to collaborate on a task. While clearly a step in the right direction, there is one disconcerting aspect of these systems: shadows. For the most part—because of the position of the projector—shadows thrown by the users are along the periphery of the table. However, when a user leans across the table to move an object, their shadow moves inward as well, and may occlude projected data. As the number of collaborators increases, the likelihood of interfering shadows increases, which may impede collaboration. Additionally, there are shadows thrown by the various components of the system itself: the *Urp* model buildings throw their own shadows, independent of the computed shadow for time of day. This may
result in two incompatible shadows, a characteristic which is distracting at best.

Other researchers embed computation inside the tangible object which is used as both the input and the output device. We refer to such TUIs as standalone devices. One example of a standalone device is the curlybot\cite{46} which is useful for teaching geometry concepts. Shaped like a half sphere with two wheels on bottom and a single record/play button on top, the student—whether pre-school age or college age—records geometric shapes by maneuvering the curlybot on a tabletop (in 2D space). The recorded motion can then be played back continuously, allowing the student to experiment with ways in which different shapes can be drawn with just part of a pattern. For example, to make a circle, only a small arc must be recorded. Upon playback, the curlybot would continuously draw the arc, resulting in a circle.\footnote{However, on a visit to MIT in 2002, I had the opportunity to use the curlybot, and found that it was too sensitive to successfully draw many patterns, including a smooth circle.} Being that this system is comprised of only a single component, it is not exactly conducive to collaborative learning, although it may be interesting to see what can be done with multiple devices.

As we continued our research, advances were made by other researchers as well. Sensetable, by Patten et al.\cite{120}, uses electromagnets to track the unspecialized components of a system on which users are able to set related parameters. To give the pieces meaning, an overhead projector was used to display details on and near the components. Although using electromagnets instead of computer vision eliminates the problem of camera occlusion, the overhead projection may still be obscured by users’ shadows.

### 4.1.2 Tangram

Our prototype system, at first glance, appears to be a standard tangram puzzle with pieces set out on a tabletop: the shapes can be moved freely about, rotated, and picked up. However, the seven puzzle pieces are each labeled with an identifying sequence of reflective colored dots which serve as fiducials (see Figure 4.3). Since like pieces are interchangeable, they are labeled identically; we have no need to differentiate between them. Rather than tracking the
pieces using an overhead camera, to minimize occlusion, our camera is mounted beneath a semi-transparent Plexiglas tabletop surface, beside a light source, with puzzle pieces placed fiducial-side down. (See Figure 4.4.) The semi-transparency of the Plexiglas effectively blocks from view any objects in the environment that are not placed on the table. In this way, the computer vision can more easily isolate the fiducial markers.

Figure 4.3: Fiducial tags on the backs of the pieces identify them. Photo courtesy of L. Scarlatos. (The white lines are added, to clearly demarcate the pieces.)

Figure 4.4: Diagram of TICLE table setup. Photo courtesy of L. Scarlatos.

After identifying the puzzle pieces and their positions, the system generates a string.\footnote{Although we do not project onto the tabletop, a semi-transparent surface can serve as an ideal projection screen for times when that capability is required.} For those not familiar with programming jargon, a string is a grouping of characters.\footnote{For those not familiar with programming jargon, a string is a grouping of characters.}
representing the state of the puzzle which is comprised of one substring per adjacency. There is no substring for pieces that do not touch. The substrings are sorted alphabetically and concatenated to produce a unique representation of the puzzle state that is translation and rotation invariant.\cite{Scarlatos} Developed by Scarlatos, Appendix B provides an abbreviated explanation of puzzle-representation strings. See Scarlatos\cite{Scarlatos1, Scarlatos2} for more detail.

By comparing this string with the solution string, we decide on appropriate feedback. If the two strings are identical, we know the current state is the solution, and offer congratulations. If one or more of the substrings comprising the solution is found in the current state’s string, we know there is a partial solution, and may offer encouragement and/or make available a related hint on the nearby computer display. The choice would depend on how the current string compares with previous ones: if a correct substring has not shown up previously, we let them know they’re making progress, but if all correct substrings have been there a while, they may be stuck. Whatever is determined, the touchscreen display presents the users with relevant feedback, such as displaying the current state of the puzzle as our software sees it, and possible actions they may take, such as reviewing the objectives of the game or receiving hints.

![Figure 4.5: Screen shot of a hint given by TICLE](image)

Acting as a guide-on-the-side by giving encouragement and hints, the computer prevents the user from getting frustrated or losing interest. Hints, which are framed as questions,
are intended to get the users thinking about how some of the pieces may be put together to form another shape. For example, “How can you make a triangle out of two smaller triangles?” or “How can you make a triangle out of two triangles and a square?” Hints have both audio and animation components to engage users, but can be ignored if the students prefer to concentrate on the learning activity alone. Although the current implementation cycles through hints in a logical order, a more intelligent version would choose the most relevant hint to present.

We now review two case studies taken from research papers that I co-authored with Scarlatos and others [145, 146], and which pertain to the TICLE tangram prototype.

### 4.1.3 Case Study 1

Ever since the Exploratorium opened its doors in San Francisco in 1969, hands-on children’s museums have proliferated across the country. Due to their constructivist nature, they are conducive to informal learning, including that of mathematics. We therefore chose to install a TICLE prototype at the Goudreau Museum of Mathematics in Art and Science (see Figure 4.6), an established hands-on children’s museum which regularly had groups of children coming, to conduct a usability study.

![Figure 4.6: TICLE tangram at the Goudreau Museum. Photos courtesy of L. Scarlatos.](image)
We were interested in answering the following questions.

- Does feedback from the computer system keep students engaged so they don’t give up so quickly?

- Do all students participate and work together?

- Do the hints stimulate metacognitive thought processes\(^8\), leading to understanding and solving the puzzle?

- Does understanding gained from the activity transfer to similar problems?

Three test sequences were conducted using 4\(^{th}\) and 5\(^{th}\) graders, with two groups of three children participating in each test sequence. In the first phase of the test, one group worked with our TICLE tangram system while the other group, the control, used a physical tangram puzzle with no supplemental guidance. Both groups were videotaped as they attempted to construct a square using all seven tangram pieces in the 15 minutes allotted for the task. In the second phase of the test, the children were asked to use the same pieces to construct the house shape, as in Figure 4.2(b) on page 64. In the third phase of the test, students were asked a set of questions that included their impressions of our prototype.

We learned much from this study. For instance, all of the groups initially tried to solve the puzzle unaided, but eventually asked for help. The hints that our system gave did seem to help: although no one solved the puzzle, those using the prototype came closer to a solution, putting more of the pieces together correctly. The hints also seemed to inspire thoughtful discussion within the groups. In general, the control groups wanted to “give up” sooner than those using the computer system. In the second phase of the test, all participants were able to construct the house shape.

The subsequent interviews were also revealing. When asked about our computer interface, two-thirds thought that the computer’s feedback was helpful. The remainder thought that

---

8Metacognition—cognition about cognition—refers to a person’s knowledge of, and regulation of, their own thought processes. It includes what one knows about oneself as a learner as well skills like planning and evaluating.
getting hints was “cheating”, or found the feedback to be distracting (Figure 4.7). When asked how the prototype could be improved, 55% suggested creating more or better hints and 22% thought that the hints should simply remain on-screen for longer (Figure 4.8).

In addition to the hint improvement suggestions, subjects’ suggestions for improving the overall prototype included making the puzzle pieces smaller or out of different materials and providing an outline to place the pieces in.

Our study suggested several improvements to the TICLE prototype, which were then implemented. These include:

- Developing additional hints that more fully reflect metacognitive dialogs (e.g., discussing why a particular approach will work, such as “If we do . . . , then we’ll have two triangles which we could put together to form a square.”), support scaffolding, and present similar concepts in different ways;

- Increasing context sensitivity (i.e., making the computer more sensitive to the state of the puzzle so that presented hints are those that are most relevant) by developing
Figure 4.8: Subjects’ recommendations for how to improve TICLE hints.

![Bar chart showing hint improvement suggestions]

- A simple rule base that is triggered by the presence, or absence, of substrings representing partial solutions in the current state;

- Supplementing the audio with text, to compensate for excessive noise in the museum and to accommodate those with hearing impairments;

- Using a touch screen (instead of a mouse) to interact with the computer (i.e. ask for hints).

In addition to prototype shortcomings, we also found that the setup of our experiment was not ideal. While some groups of three children work well together, we found that the group will often break into two subgroups, whether the lone child is working with the system, having taken control, or the two-person group leaving out a third child. This seemed to show that two people may be the ideal group size for this particular application. Also, the allocated

---

9A rule base is a list of rules pertaining to the system, e.g., if substring X, representing the two large triangles are placed side-by-side to form a larger triangle, occurs in the current-state string—half of the solution—the system should display a hint about building the second half of the square.
15 minutes seemed to be too long for most children to maintain interest, and despite our assurances that they did not have to spend the full time working at the puzzle, many felt obligated.

### 4.1.4 Case Study 2

In addition to the improvements to the TICLE prototype that were made after our case study, we also realized that another study was in order, incorporating the use of more rigorous testing methods. In the second study we adjusted some parameters based on our findings in the first study, such as the number of people in each group (two instead of three) and allocated time (10 minutes instead of 15). We also adjusted our test assumptions to the following:

- A “guide on the side” providing hints, encouragement, and reminders about the objectives will help to motivate students and keep them from getting distracted or giving up too soon.

- Context-sensitive hints will get students to think about the problem in new ways, leading to more metacognitive activities.
Because they are focused and thinking more deeply about the problem, groups using TICLE will be more likely to solve the puzzle. (See Figure 4.10 on page 77.)

Working with a TICLE system will help children to develop problem-solving skills that will transfer to similar problems.

4.1.4.1 Case Study Logistics

We ran two separate test sessions at the Goudreau Museum: one with a group of girl scouts finishing third grade, and the other with a group of boy scouts finishing second grade. In each session, we divided the students into six groups of two. Groups were assigned based on where the students were sitting: we assumed that children who sat next to each other would work well enough together as a team. Children wore labels marked with consecutive numbers, which were used to identify them in our observations.

The test itself was conducted in three parts:

1. **Each team was asked to make a square using all of the Tangram pieces.**

   Half of the teams got to use the TICLE system for this part. The other half used the museum’s regular Tangram puzzle, with no computer or teacher aid, but with an instruction sheet. Video cameras were stationed at both the TICLE table and the regular table, to record student actions. The children were told that they had up to ten minutes in which to solve the problem; they were able to quit sooner if they wanted.

2. **The teams were asked to make a rectangle using the Tangram pieces.** All teams used regular Tangram pieces at a separate set of tables. Video cameras were also stationed at these tables to record student actions. Once again, they had up to ten minutes to solve the problem.

3. **Team members were interviewed individually.** They were asked a fixed set of questions regarding their team, including how they felt about collaboration; the Tangram, and what they thought they learned from it; and the TICLE interface,
including pros, cons, and suggestions for improvement. These interviews were recorded on audio tape. Interview questions are included as Appendix C.

4.1.4.2 Evaluation Methodology

For evaluation purposes, we adopted Artzt and Armour-Thomas’s cognitive-metacognitive framework for protocol analysis[5]. This framework was designed to “differentiate explicitly between cognitive and metacognitive problem solving behaviors observed within the different episodes of problem solving.” Metacognition, which is basically “cognition about cognition”, includes understanding, analysis, planning, and verification: essential steps in mathematical problem-solving, according to Polya[125]. Alternatively, cognition may be observed in behaviors such as watching, listening, exploring, and implementing a plan.

Like Artzt and Armour-Thomas, we used our videotapes of the tests to classify the observable behaviors of the teams. We produced a table with rows representing the behaviors: understanding, analyzing, exploring, planning, implementing, verifying, watching/listening, and distracted/fooling-around. We added this last category (essentially a non-cognitive behavior) to the framework because it was something that seemed to happen a lot. Columns in our table represented time slices of thirty seconds. With a separate table for each team, we marked all of the observed behaviors for each time slice, using letters (rather than the children’s assigned identification numbers, for clarity) to indicate who was doing what. Because we couldn’t know what the children were thinking, we only recorded metacognitive behaviors when the children’s comments suggested that they were thinking about the problem in a particular way. We also marked the tables with other significant events such as children’s comments and discussions, getting hints, and stealing a teammate’s hat.

Table 4.1 shows a table for a sample team of girl scouts that worked particularly well together. They frequently stood back to analyze the situation. For them, the hints caused them to think about the problem in new ways that ultimately led to the solution. The top
row in this table labels the time slices in which the behaviors were observed. The numbers beneath the table correspond to observable events that are described in the text.

<table>
<thead>
<tr>
<th>Time Interval</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meta-cognitive</td>
<td>understand</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>analyze</td>
<td>B</td>
<td>A</td>
<td></td>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>verify</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>plan</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive</td>
<td>explore</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>implement</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>watch/listen</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>Non-cognitive</td>
<td>fooling</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>around</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.1: Observed behaviors for sample team

1. The subjects start by making a triangle (half of the square) using the 2 large triangles.
   We see this as a great start.

2. B: “That’s going to pop out.”

3. The subjects watch a hint.

4. A: “Try to form a triangle” using the remaining pieces.

5. The subjects watch a hint.

6. B: “Let me try something ... I got it!” Subject B starts building the another triangle on top of the triangles referred to in comment (1).
Figure 4.10: Number of teams that solved the 1st puzzle (with and without TICLE), and number of teams that solved the 2nd puzzle (conventional puzzle only). All values are out of 6 teams.

7. A: “But what about that space?”

8. A: “I got an idea.”

9. Finish building 2nd half of the square on top of the 1st half. A: “Now we have to transport that there.”


4.1.4.3 Observations

Our observations strongly suggest that the TICLE system does indeed keep students focused and help them to solve the puzzle. Figure 4.10 shows that four out of six teams using TICLE were able to make a square from the pieces; only one out of six teams using a conventional tangram were able to do that.

Table 4.2 shows a summary of our observations of the first part of the test. For each team, the table shows the percentage of time slices spent on metacognitive behaviors (understanding, analyzing, planning, verifying), cognitive behaviors (exploring,
implementing, watching, and listening), and non-cognitive behaviors (fooling around or distracted). Because several different behaviors (of both team members) may have been observed within a single time slice, the sum of the percentages often exceeds 100%.\(^{10}\) This table also shows which teams solved the first problem (making a square using TICLE or a conventional tangram) and the second problem (making a rectangle using conventional tangram pieces). Teams labeled with a “G” are comprised of girl scouts; teams labeled with a “B” are comprised of boy scouts.

<table>
<thead>
<tr>
<th>Team</th>
<th>Metacognitive</th>
<th>Cognitive</th>
<th>Fooling</th>
<th>Solve #1</th>
<th>Solve #2</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1</td>
<td>56%</td>
<td>100%</td>
<td>0%</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>G3</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td>✓</td>
<td>×</td>
</tr>
<tr>
<td>G5</td>
<td>17%</td>
<td>100%</td>
<td>0%</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>B1</td>
<td>29%</td>
<td>100%</td>
<td>0%</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>B3</td>
<td>0%</td>
<td>100%</td>
<td>5%</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>B5</td>
<td>10%</td>
<td>100%</td>
<td>30%</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>G2</td>
<td>0%</td>
<td>100%</td>
<td>67%</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>G4</td>
<td>6%</td>
<td>100%</td>
<td>31%</td>
<td>×</td>
<td>✓</td>
</tr>
<tr>
<td>G6</td>
<td>0%</td>
<td>77%</td>
<td>92%</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>B2</td>
<td>20%</td>
<td>100%</td>
<td>0%</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>B4</td>
<td>20%</td>
<td>100%</td>
<td>55%</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>B6</td>
<td>11%</td>
<td>100%</td>
<td>72%</td>
<td>×</td>
<td>×</td>
</tr>
</tbody>
</table>

Table 4.2: Summary of observations of the first part of the usability test. (In the last two columns, ✓ indicates the team solved the puzzle, while × indicates they did not.)

Also encouraging is that our observations suggest that the lessons learned from using TICLE may transfer to similar problems. Figure 4.10 shows how many teams were able to construct the rectangle using a conventional tangram, and Table 4.2’s last column show which teams these were. Three of the six teams (50%) that had used TICLE earlier were able to solve the problem (G1, G5, B1). Those three had also completed the square (second to last column of the table). The one team that had solved the first problem, but not the second (G3), had solved it very quickly with no discussion and without looking at any hints. Perhaps

\(^{10}\)This is so for all teams except G3.
one of the team members had seen the puzzle before, and simply remembered the solution. On the other hand, only two of the teams that had used a conventional tangram earlier were able to construct the rectangle (G2, G4). One of those teams had also successfully constructed the square (G2).

As shown in the table, teams using TICLE tended to have more metacognitive interchanges. Many times when the children would get a hint, we would hear “Oh, yeah” and see a flurry of activity indicating that the hint had caused them to think about the problem in a new way. Sometimes the hints helped in unexpected ways. For example, when one group was told that pieces shouldn’t be stacked on top of one another (because the system had lost track of a piece), they got the idea of building the second half of the square on top of the first half of the square, and then moving it into place. For them, this was the key to the solution (which they ultimately found).

The table also shows that TICLE users spent a lot less time fooling around. We noticed that in several of the teams, the children would take turns working on the puzzle. For those teams using TICLE, the child who was not exploring with the puzzle would often be looking at the hints. For the control teams, other goings-on in the museum proved far too tempting for the idle teammate. We also observed a few of the control teams making comments such as “Can you help us?”, “This is impossible!”, and “Can we use the computer now?”. No one using TICLE made such comments.

Children in the control groups were more frustrated, first of all, because they had not computer to help them, but also because they saw that other children did have computer help which was giving them valuable hints. They may have been upset that the other teams were making progress with the aid of the computer while they were denied it. This is something to think about in setting up future experiments.

Interviews at the end of the study once again showed that our system could be improved. As in our first case study, several children who used the TICLE system said that the voice

---

11 This is supported by the lack of any metacognitive or non-cognitive events for this team.
offering encouragement was “annoying” and “distracting”. In addition, a few of the children thought that we needed to create more hints.

4.1.4.4 Statistical Analyses

Since the data we collected was categorical, the non-parametric $\chi^2$ test for homogeneity was used to analyze the data, with the null hypothesis $H_0$ being that there is no difference between using TICLE and not using TICLE, in the number of groups that complete the tangram, and the alternative hypothesis $H_1$ being that there is indeed a difference. Our data is given in the $2 \times 2$ contingency table seen in Table 4.3.

<table>
<thead>
<tr>
<th></th>
<th>Solved</th>
<th>Not Solved</th>
</tr>
</thead>
<tbody>
<tr>
<td>With TICLE</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Without TICLE</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 4.3: Expected values (left) and observed values (right) for the $\chi^2$ analysis of the solving of the tangram. The table on the right is the contingency table.

We then use the formula $\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$, where $k$ is the number of cells in the table (4), $E_i$ is the expected value, and $O_i$ is the observed value. The result, $10/3$ or $3.33$, would typically be looked up in the $\chi^2$ table in the $0.05$ column, one degree-of-freedom row. However, since the alternative hypothesis is directional and uses a $2 \times 2$ contingency table, to evaluate it “at the .05 level, the appropriate critical value to employ is $\chi^2_{90}$, which is the tabled chi-square value at the .10 level of significance.” Thus, since $3.33 \geq 2.71$, we are able to reject the null hypothesis, and accept the alternative hypothesis: TICLE did have a statistically significant impact on the number of groups able to solve the tangram.

To see whether TICLE had a statistically significant impact on children’s focus on the puzzle, the Mann-Whitney $U$ test was employed, with the teams ranked by how much

---

120.05 is typically used as the cut-off for significance.
13Degrees of freedom = (number of rows - 1)(number of columns - 1)
14Testing for a relationship in only one direction (one-tailed).
15Although some sources are against using the $\chi^2$ test for very small sample sizes, both the Yates’ correction for continuity and Fisher exact test give overly conservative results.
16Being that 1) the data was extracted from video by the researchers rather than by blind assessors and
fooling around they did: a lower rank meant less fooling around and a higher rank meant more. The null hypothesis \( H_0 \) is that TICLE did not have an effect on their focus, i.e., the sum of the ranks of teams using TICLE is equivalent to the sum of the ranks of the teams not using TICLE. The alternative hypothesis \( H_1 \) is that TICLE did affect the focus of the students, i.e., the sum of the ranks of teams using TICLE is less than the sum of the ranks of teams not using TICLE. The rankings for the teams are seen in Table 4.4 with rank adjustment for tied scores.

<table>
<thead>
<tr>
<th>Team</th>
<th>G1</th>
<th>G3</th>
<th>G5</th>
<th>B1</th>
<th>B2</th>
<th>B3</th>
<th>B5</th>
<th>G4</th>
<th>B4</th>
<th>G2</th>
<th>B6</th>
<th>G6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>xT</td>
<td>T</td>
<td>T</td>
<td>xT</td>
<td>xT</td>
<td>xT</td>
<td>xT</td>
<td>xT</td>
</tr>
<tr>
<td>Fooling around</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>30</td>
<td>31</td>
<td>55</td>
<td>67</td>
<td>72</td>
<td>92</td>
</tr>
<tr>
<td>Rank prior to tie adjustment</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>Tie-adjusted rank</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
</tr>
</tbody>
</table>

Table 4.4: Rankings for Mann-Whitney \( U \) test. Condition ‘T’ means the group did use TICLE, ‘xT’ means they did not. Fooling around given in percentage of time slices with fooling around observed.

The sum of the tied ranks for the TICLE condition (\( \Sigma R_1 \)) is 25 and the sum of the tied ranks for the no TICLE condition (\( \Sigma R_2 \)) is 53. \( U_1 \) and \( U_2 \) are computed from the following equations: \( U_1 = n_1 n_2 + \frac{n_1 (n_1 + 1)}{2} - \Sigma R_1 \) and \( U_2 = n_1 n_2 + \frac{n_2 (n_2 + 1)}{2} - \Sigma R_2 \), where \( n_1 \) is the number of samples in the experimental group, and \( n_2 \) is the number of samples in the control group. The values we get are \( U_1 = 32 \) and \( U_2 = 4 \). The smaller of the two values, in this case 4, is the \( U \) statistic which is looked up in the table of critical values for the Mann-Whitney \( U \) test. For \( n_1 = 6 \) and \( n_2 = 6 \), the critical one-tailed 0.05 value is \( U_{0.05} = 7 \). Since, to be significant, the obtained \( U \) value must be less than or equal to the tabled critical value, we find that we have significance at the 0.05 level, and can reject the null hypothesis. It is

---

2) the data is in percentages of 30-second time slices that contained fooling around, not absolute time, it should be treated as ordinal data, not continuous. Therefore, the Mann-Whitney \( U \) test was used rather than the \( t \)-test.

17 This, too, is a one-tailed hypothesis.

18 At the 0.01 level, the obtained \( U \) value must be less than or equal to \( U_{0.01} = 3 \), which is not the case. However, since transforming data into ranks sacrifices information, making it less likely to find significance, and since the difference between values is quite large, it is possible that the results are
clear from this that TICLE helped keep the children focused on the task at hand.\footnote{Additional support for this is the request for more hints; the children themselves believed that the hints were helping.}

### 4.1.5 Conclusions

The results of our case studies are encouraging: children using our TICLE system were more likely to solve the presented problem than those using a conventional puzzle with no assistance. Furthermore, our observations suggest that children that use TICLE may be better able to solve related problems, even when they are given no other assistance. As we saw in Table \ref{table:4.2}, 80\% of teams that solved the tangram as a square went on to solve it as a rectangle and we believe that they will be able to solve other related problems as well.

Our observations also seem to support our other assumptions. First, it is now evident that a computer “guide-on-the-side” can help to motivate students and keep them from getting distracted or giving up too soon. Second, context-sensitive hints do seem to get students to think about the problem in new ways. We’ve seen that they do lead to more fruitful discussions and actions, suggesting that they trigger more metacognitive activities. The implications are clear. Tangible interfaces do indeed provide a new way for us to enhance our children’s education without forcing them to sit in front of a computer. Instead of being the focus of educational activities, the computer can now take on a new role: guide on the side. The possibilities are endless.

### 4.2 Three-Dimensional Tangible User Interfaces

As the tangram is inherently a two-dimensional puzzle, the next phase of this project was to build on what we had learned in the first phase, and extend TICLE to the third dimension, tracking 3D objects within a closed environment. The National Council of Teachers of Mathematics (NCTM) website on Principles and Standards for School Mathematics

\footnote{significant at a higher level.}
Figure 4.11: Soma cube: (a) pieces and (b) solved as a cube.

[standards.nctm.org] states that “Instructional programs from prekindergarten through grade 12 should enable all students to use visualization, spatial reasoning, and geometric modeling to solve problems.” For grades 9–12, this means students should be able to “use geometric models to gain insights into, and answer questions in, other areas of mathematics” as well as “use geometric ideas to solve problems in, and gain insights into, other disciplines and other areas of interest such as art and architecture.”

To this end, we chose to implement a Soma cube puzzle (see Figure 4.11), which, according to Weisstein[186] is “[a] solid dissection puzzle . . . There are seven soma pieces composed of all irregular face-joined cubes (polycubes) with \( \leq 4 \) cubes. The object is to assemble the pieces into a cube. There are 240 essentially distinct ways of doing so.”

Such a puzzle combined with sensor technology and a simple guide-on-the-side program, gives students the means to explore relationships among geometrically similar objects.

As we discussed earlier, it is nearly impossible to implement 3D object tracking using computer vision, so the need arose to find a different approach. Because of the complex structure of the Soma cube pieces, we decided to investigate technologies by using them on simple cubes. We chose Instant Insanity, played with four cubes, as a game to implement in the interim. Using four different colors to paint the six faces of the four cubes, “The object of the game is to place the cubes in a column of four such that all four (different) colors
appear on each of the four sides of the column.” Details on how to solve Instant Insanity can be found in Appendix A.

4.2.1 Related Work

Not many tangible user interfaces have been created which are three-dimensional by design and function. To qualify as such, component parts would be required to react in three dimensions. It is not enough for objects to have a three-dimensional shape or even to be used in what appears to be three dimensions. As an example, let’s look at Resnick et al.’s BitBall[131], a standalone TUI that can help students learn about gravity. Containing an accelerometer and colored LEDs, it can be programmed to act on the acceleration or deceleration of the ball. While we may think of a ball as a three-dimensional object operating in three dimensions, typical usage of a ball is in a single dimension—up or down—unless it is thrown with a curve, in which case it would be operating in two dimensions.

Escape Machine[189], an educational PacMan-like game in which children manipulate a tangible state machine to control the behavior of characters in a maze, is a true 3D TUI. Using the researchers’ own Posey, a computationally-enhanced hub-and-strut construction kit, to build the state machine, 9–11 year olds demonstrated that they were capable not only of manipulating the tangible state machine to complete the game, but also at times “formulate[d] higher-level strategies for play.” Posey’s hubs contain an array of photosensors which “see” the infrared LEDs contained within the strut connected to it.[188] Yaw, pitch, and roll of the strut within the hub can be computed from the data.

Smart Blocks[51] are a set of blocks aimed at “facilitating hands-on learning of the volume and surface area of 3D shapes.” Blocks can be connected to each other to form shapes that are recognized by the system, which then calculates the volume and surface area of the shape. Multiple users can manipulate the blocks simultaneously, and feedback is provided on a nearby display. Two modes exist, one for trial and error exploration, and one for testing.

---

20 Rotation about the $z$, $y$, and $x$ axes respectively.
While this system may be promising, it suffers from a number of drawbacks. Each of the six faces of each block contains a hole in its center into which a dowel connector is inserted to connect it to another block. This requirement not only interferes with how the users interact with the blocks, but affects the possible shapes that can be assembled: attaching a block with a single connector can be done, but adding a block requiring two or more connectors is physically impossible. Additionally, the RFID technology that was used allows identification of present pieces, but the system has no way to interpret the shape of the construction, or even verify that a single shape has been constructed. Nothing prevents a user from gathering some blocks and some connectors and merely placing them together on the workspace.

ActiveCube[91] [184] is a set of small cubes that connect to each other in any arrangement, allowing users to construct 3D shapes which they can use to interact in 3D environments. Working in real-time, the computer recognizes the structure and adjusts the on-screen representation, much as we did with the tangram. Additional sensors can be incorporated into the various cubes to give the structure functionality. Although these cubes work well for building virtual structures, the faces of the cubes are rotationally-equivalent and there is no way to note rotation. Another drawback of these cubes is that they are connected to the computer by a wire link to a base cube; this restricts users to working around that cube, physically and conceptually.

Although computer vision is typically not suitable for tracking three-dimensional objects, Baudisch et al.[12] have devised a way to do it for their Lumino blocks, using fiber optic bundles embedded in the objects. The blocks can be manipulated and stacked on a tabletop. Allowing light to pass through the blocks permits the fiducial markers of objects above to be seen, and join with the fiducial of lower blocks to form a unique tag which defines how pieces are situated. However, there are many limitations, including a decrease in resolution as the number of objects stacked increases, and an inability to freely rotate objects about all of the axis.\footnote{In most cases, objects can only be freely rotated about the y-axis, but in some cases they can be rotated 180° about the other axes.}
Some tangible user interfaces show promise for use not just in general education, but also in special education. Topobo\textsuperscript{[118, 129]}, is a standalone TUI device which allows youngsters to construct their own animal-like robots and teach them, by example, how to move. Another hub-and-strut construction kit, Topobo’s active hubs record movement of attached strut limbs and then play back this motion. Raffe et al.\textsuperscript{129} believes that the ability of eighth graders to build a variety of robots capable of walking indicates that Topobo “can support understanding how balance, leverage and gravity affect moving structures.” Parkes et al.\textsuperscript{118} observed educators working with Topobo in a classroom setting, and discovered that for one teacher working with a group of 8–14 years olds with special needs, including some with ADHD and Asperger’s syndrome, “Topobo kept them very focused but . . . they needed directed and guided tasks, such as small specific problems to solve or very detailed instructions to follow.”

4.2.2 Implementation

Our goal was to find technology to use to track the four cubes of Instant Insanity, which we would then extrapolate for use in the Soma cube pieces and combine that with a hint module\textsuperscript{22} as we did with the tangram. Instant Insanity, also known as “The Great Tantalizer,” consists of four cubes, with the color arrangements as shown in Figure 4.12 but with any color substitutions.

Although there are $4! \times 244$ possible arrangements of the cubes—they can be stacked in any order, and each cube has 24 orientations—the order of the cubes does not matter, and any solution would have eight orientations, for a total of $24^4/8$ \textit{(41,472)} distinct arrangements. Of these, there is a unique solution to the problem. A similar puzzle uses clocks instead of colors—2 o’clock replaces yellow, 4 o’clock replaces green, 8 o’clock replaces blue, and 10 o’clock replaces red. Arranging the cubes such that each side has one of each color would be

\textsuperscript{22}There are no good hints for Instant Insanity aside from suggesting the users rotate the pieces or check to make sure there is one of each color per side, if they have not reached a solution; mostly this is done by trial and error or with a graph as discussed in Appendix A.
equivalent to the original Instant Insanity. For an easier variation, the cubes can be arranged such that each side sums to 24, which can be done in three ways: one each of 2, 4, 8, and 10; or 2, 2, 10, and 10; or 4, 4, 8, and 8.

Because the original Instant Insanity arrangement of colors has just one solution, I decided to rearrange the color configuration to make it easier to solve. This color configuration also allows the pieces to be put together such that each side is entirely one color—impossible with the original coloring. The spray paint colors we were able to obtain to produce the cubes were slightly different than those of the original Instant Insanity, as seen in Figure 4.13.

In addition, each side will have one of the numbers 2, 4, 8, or 10, to allow for a wider range of problems. Each color will not in any way be related to the numerical value on it; they will be taken individually (see Figure 4.14). In fact, the number arrangement corresponds to the
colors on the original Instant Insanity cubes so that, if asked to find a numerical solution using distinct numbers on each side rather than any arrangement of a sum of 24, it becomes the original Instant Insanity problem.\footnote{Figure 4.12's colors red, yellow, green, and blue are equivalent to Figure 4.14's numbers 10, 2, 4, and 8, in that order.}

The problem to solve with the puzzle cubes would be displayed on the screen as well as spoken. The user will be able to click on an audio button to have the problem repeated or a help button for a hint. The computer acts as a guide-on-the-side, always keeping track of what the user is doing by checking on the state of the puzzle. If the user needs help, the computer will give a hint, but not the answer. When the user successfully solves the problem, the computer “pats” the student on the back, and moves on to another problem.

The problems may include the following: “Put the cubes together so that each side is one color—one side red, one side blue, one side yellow, one side green,” or “Put the cubes together so that each side sums to 24.”

Hints would be designed to make the user think about the problem rather than telling them what to do, and may include, “Pick a color for one of the sides. Work from there,” “Look for a pattern of colors that is the same on all pieces,” or “How many ways can you use the numbers 2, 4, 8, and 10 to sum to 24?” Helpful reminders will also be given, when deemed necessary, such as, “Make sure the four pieces form a row” or “All four cubes must
be used in the solution.”

4.2.3 Technology

Self-contained technology that would recognize when two disconnected objects (in this case cubes) are touching each other and how they are touching was needed, along with a means to wirelessly communicate this information back to the computer. We decided to use Basic Stamps with a radio frequency transceiver\textsuperscript{24} within each cube, wired to two “active leads” on each of the six faces (see Figure 4.15 for a diagram representing connection to a single face). In Figure 4.16, these active leads are represented by the solid unfilled squares. The dashed squares represent passive leads, wired to the opposite passive lead on the same cube face, by way of a unique resistor, acting as an ID. In addition to the RF transceiver, each Basic Stamp has one capacitor for each side, which acts as a sensor. Since the resistance on any possible circuit is different, the amount of time it takes to charge the capacitor varies depending on the circuit that is completed. It is this length of time that is being used to differentiate among circuits.

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{sensor_diagram.png}
\caption{Sensor diagram}
\end{figure}

Although the two sets of passive leads per side should enable the computer to differentiate between $0^\circ/180^\circ$ and $90^\circ/270^\circ$ rotations, an extra set of leads enables it to differentiate between $0^\circ$ and $180^\circ$ and between $90^\circ$ and $270^\circ$ rotations (the filled squares in fig 4.16).

\textsuperscript{24}Transceivers can transmit as well as receive information. We used the receiving capability for synchronization.
These create an additional circuit that is closed with a rotation of $0^\circ$ or $90^\circ$, but not $180^\circ$ or $270^\circ$. This connection will only be tested if it has been determined from the primary leads that two cube faces are touching. Figure 4.17 shows our prototype cubes with all leads. The numbers were moved to the corners so that they would not be blocked by the leads, and are placed so that they are readable from any direction.
enum { RED, YELLOW, GREEN, BLUE };
Piece {
    int pieceID;
    Side side[6];
}
Side {
    int sideID, color, val;
    Side *top, *right,     // sides of current piece
        *bottom, *left;
    Piece *piecelink;     // touching which piece
    Side *sidelink;       // side of the touching piece
    int rotation;         // 0, 90, 180, or 270
}

Listing 4.1: A possible data structure to represent the pieces.

completely, which in the case of Instant Insanity is part of the problem definition\footnote{26}. When an adjacency occurs, circuits are completed inside each cube (this redundancy can be used for error-checking).

The computer is hooked up to an RF transceiver as well, ready to receive data, and continuously polls all of the cubes for their adjacencies (time to charge each of the six capacitors) in a round robin fashion. The computer then takes this data and forms an internal image of the configuration of the puzzle.

### 4.2.4 Software Representation

Although the plan was to use an internal representation of the 3D puzzle similar to the one we used to represent the state of the 2D tangram puzzle (see Scarlatos et al.\cite{148}) for the Soma cube, since we were no longer concerned with 3D construction, but rather just adjacencies, we were able to simplify the representation.

To determine the state of the puzzle, some information must be stored for each piece, as shown in Listing\ref{1.1}

Each piece stores its own identification number, the IDs of the pieces adjacent to it (or

\footnote{26This is true for the Soma cube as well, but see \ref{4.2.6} on further research for more discussion.}
links to them), the color on each side, the value on each side and the four sides of that piece each side is touching. All of this information must be entered beforehand. The links are stored to expedite testing of the solution. The side with the lowest ID number will be the “head” (see Figure 4.18). The next lowest ID number, which should be touching the head is “side 1”, followed by the next lowest ID number, which is “side 2” and also touches the head, but is not opposite from side 1. Testing the solution will be done by starting at the head of each piece and moving through all of the sides of the pieces in order of ID. Each piece must be touching on either one or two sides—two of them on one side (end pieces), two on two (opposite) sides. The computer then constructs an image of the configuration of the puzzle based on the adjacencies. The state of the puzzle is stored in the links within the pieces. Testing for a solution is now straightforward: look at each side of the puzzle and check for compliance with the stated problem. The simplified algorithm for the problem of having each color appear on each side is shown in Algorithm 1.

4.2.5 Evaluation

The original plan, called for testing this puzzle as follows:

To test whether playing with this puzzle helps students learn the relationship
Algorithm 1 Algorithm for testing for a solution.

\begin{algorithm}
\begin{algorithmic}
\STATE start at an end piece
\STATE find side with adjacency
\FOR{$i$ in top, right, bottom, left}
\STATE find color of side[$i$], mark it as seen
\STATE $k \leftarrow i$
\WHILE{more pieces}
\STATE follow $k$-link to next piece $j$ in top, right, ...
\IF{$j$-color already marked}
\STATE not a solution, return
\ENDIF
\STATE $k \leftarrow j$
\ENDWHILE
\ENDFOR
\STATE solution found
\end{algorithmic}
\end{algorithm}

among numbers, I will conduct the following experiment on groups of third-graders. First, the group will be divided into two—a control group and experimental group. Both groups will begin with a short, timed assessment on the topic. Those in the experimental group will subsequently use the experimental system, while the others will be given illustrated instructions on a poster. Both groups will be allowed a set maximum amount of time to play with the puzzle. Finally, both groups will end with another timed test to see what affect playing with the puzzle had on their understanding of the subject matter.

After the post-test, I plan on giving both groups an explanation of the problem and showing them ways of representing the problem to make it easier to come to the solution. Although it might make sense to make this explanation part of the system, I would like to test the impact of the puzzle itself. If I were to give an explanation before the post-test, it would have to be to both groups, and then I wouldn’t know if the students learned from the puzzle or from the explanation. Also, if I would want to take a truly hands-off approach, I would need to leave a written booklet for the control group rather than give a short presentation on a computer after playing with the puzzle, as I would for the experimental group.
This would be an unfair advantage for the control group given that they would have access to it the whole time.

I will look at the technique as well as overall time required to complete the tests. I believe I will see a substantial score increase for those in the experimental group. In addition, I will compare the amount of time students in each group play with the puzzle before they get bored. I expect the experimental group to remain involved for a longer time.

Due to the issues encountered over the course of this phase of the project, we did not do any testing. When these issues are resolved\textsuperscript{27}, the described evaluation should be carried out.

4.2.6 Further Research

Belatedly, it was discovered that working with resistors was not as straightforward as it was believed it would be, and the project suffered the same problem of dealing with inconsistent resistance values as Camarata et al.\textsuperscript{22} did in the making of a sensor. Although in actuality the successful tracking of the Instant Insanity cubes was not implemented, on paper, a clever approach to differentiate among cube rotation was designed, which \textit{should} work with reliable technology. This design should work not only with cubes, but with objects comprised of cubes, such as the Soma cube pieces. Treating the individual cubes as separate, albeit permanently connected, cubes, is one possible approach, but contains unnecessary circuitry for sensing what we \textit{know} is an adjacency. The design would best be tweaked for efficient tracking.

Building a representation of the state of the Soma cube puzzle was the plan, but upon further inspection, it appears that checking for a solution to the cube problem is anticlimactic: because the surface area of the constructed shape is minimized when in the shape of a (3x3x3) \textit{cube},

\textsuperscript{27}As I truly believe they can be.
cubeto 54 square units—it would be simple enough to tally the faces that are touching other faces. If the number of faces not touching another face is 54, we know the cube solution has been found. For other shapes with the Soma pieces, or if partial solutions are to be recognized by the computer, the additional information would be necessary.

Unlike the tangram, solutions to most Soma forms can be constructed in many ways—240 distinct solutions just for the cube! Therefore, instead of storing a string representing the precise layout of the pieces, it would be easier to store a solution-shape graph against which the current state of the puzzle can be compared.

For testing whether a given configuration of pieces solves the Instant Insanity or Soma cube problem, knowledge of the overall rotation of the combined cubes is not strictly necessary. However, an intelligent tutoring component of either of these two puzzles would be lacking without this capability. To display the current state of the puzzle as the user sees it—as we have done with the tangram—knowledge of the overall puzzle rotation is a must. Although we do not know the exact rotation of the two-dimensional tangram on the table, all users see it from a different perspective depending on where they are standing. Fortunately, doing a mental rotation of a 2D object is relatively simple. However, this is not the case with three-dimensional objects, which can be quite difficult for some to rotate mentally.

For clarity, computer images reflecting the state of a puzzle should show it right-side-up, as the puzzle exists in physical space. A gyroscope, or some other sensor, would need to be incorporated into the cubes for this. Although a single sensor would be sufficient for a completed puzzle, for partial solution display, each piece must be equipped with a sensor.

As mentioned earlier, the only kind of adjacency our sensors note is two cubes overlapping each other entirely. While this is precisely what is needed for both Instant Insanity and the Soma cube, the design would have to be reworked to support partially overlapping adjacencies.

---

28For three-dimensional orthogonal shapes, the cube is the shape with the least surface area. If the polycubes' component cubes are one cubic unit, for a total volume of 27 cubic units, the shape with smallest surface area would be a 3x3x3 cube. (There is the possibility that the polycubes do not fit into a cube—such as would be the case if a polycube were comprised of four cubes all in a row—but that is not the case here.)
Due to the stumbling blocks encountered with inconsistent resistor readings, I redirected my focus to another kind of tracking: that of people within a virtual environment.
Chapter 5. Tracking Collaborative
Gesturing within a CAVE\textsuperscript{TM}

Tracking gestures among multiple co-located users is a second approach to enhancing an environment with computers to support collaborative learning. In Chapter 4, we tracked objects in an environment; now we track the people themselves as they move about within an environment.

As discussed in the Introduction, movement is intertwined with social interaction and has documented benefits on education. A computer-supported educational environment that encourages collaboration and movement brings together some of the most effective educational tactics to form a comprehensive system that can be customized for a wide range of educational applications, with emphasis on special education.

The specific environment we are working in is a partial CAVE\textsuperscript{TM}, with dual 8-foot by 10-foot displays\textsuperscript{1} to give a sensation of partial immersion. In this two-stage project, we first detect users’ two-dimensional position on the floor for a simple collaborative art gallery application and then combine that data with the three-dimensional location of the users’ body parts (wrist and shoulder) to build a more complex application which relies on users’ moving about and gesturing, working toward a shared goal.

\textsuperscript{1}One serves as a wall in the environment, the other as the floor.
5.1 Related Work

Early research investigating virtual reality in education typically focused on solitary usage. ScienceSpace\cite{35} is a collection of three worlds created to help students explore difficult physics concepts, such as the relationship among mass, velocity, and energy, while noting the impact of gravity and friction on them. Using a head-mounted display and wearing tracking devices on the head and both hands, students use their virtual hand to select menu items and navigate.

Cruz-Neira et al.\cite{33} point out that, “One of the most important aspects of visualization is communication. For virtual reality to become an effective and complete visualization tool, it must permit more than one user in the same environment.” Although responses to the ScienceSpace worlds have been primarily enthusiastic, they are inherently individual experiences: the collaborative aspect of learning has been neglected.

In the Asteroid World\cite{106}, a pair of students collaborate while learning that the earth is a sphere. Although they work together toward a goal, they are in separate physical locations, one as an astronaut within a CAVE\textsuperscript{TM} moving about on a virtual asteroid, and the other at an ImmersaDesk, a smaller and less immersive version of the CAVE, which runs a mission control simulation.

The NICE project\cite{135, 136} presents a CAVE\textsuperscript{TM} virtual garden for children within which they can plant vegetables, explore, and meet other children. Available around the environment are “genies” to guide and provide feedback to the children. Although any number of children that will fit inside the CAVE\textsuperscript{TM} may join in, they are collectively represented by a single avatar within the virtual space. This avatar is controlled by a single child in the group who is wearing tracking devices, including a (rather unfortunately, tethered) wand that represents and controls the avatar’s hand. This joystick-like apparatus, used for navigation as well as picking up items, corresponds to pointing and clicking in 2D environments.\footnote{Special gloves, such as those used by Pierce et al.\cite{123} may preclude the use of explicit input devices.} Multiple remote CAVE\textsuperscript{TM}s running identical software may network together...
and appear within the same virtual garden, each represented by its own avatar, thereby providing some level of collaborative ability.

C-OLiVE\textsuperscript{[4]} is a VE that has middle-schoolers setting up an olive oil factory. They must troubleshoot the machinery and do various tasks, some of which require collaboration. The application is projected onto a single wall, across from which sit the users while interacting with the system with individual Xbox\textregistered controllers, much like a multi-user video game. With this setup, the researcher did not find that “interactivity and social play” affected learning, although users did feel that it positively affected the learning experience.

The Virtual Playground\textsuperscript{[139]} teaches students about fractions while laying out a playground to a set of given specifications. A wireless head tracker worn by the user provides position and orientation data within a CAVE-like environment, while a pair of active stereo glasses provides stereoscopic viewing and a wireless wand enables navigation and manipulation of virtual objects. While most photos in this paper show a solitary user, one photo seems to show two children using the application together, but no mention of this was found in the paper itself, and the complexity of displaying location-specific stereoscopic images for each user makes it unlikely.

In socio-ec(h)o\textsuperscript{[182, 183]} we see an example of what the authors call “ambient intelligence computing,” which, they state “is the embedding of computer technologies and sensors in architectural environments that combined with artificial intelligence, respond to and reason about human actions and behaviours within the environment.” In socio-ec(h)o, this boils down to using a CAVE\textsuperscript{TM} environment coupled with the Vicon motion capture system to attribute intelligence to the ambient environment. Over the seven game levels that comprise socio-ec(h)o, a group of four players try to uncover the meaning of a word clue that is presented to them. In the authors’ words, “Each level is completed when the players achieve a certain combination of body movements and positions.” As the players move towards completion of the level, the environment changes—the sounds and colors become more intense. Each player is labeled with a different configuration of five reflective markers placed
on their back. The Vicon motion capture system tracks the 3D positions of these markers in real-time. The $x$ and $y$ coordinates give the users’ locations on the floor, while the $z$ coordinate indicates whether they are standing or crouching.

*Flashlight Jigsaw*[^23] is a jigsaw-type game played on a large-screen projected display by up to three players holding spatially tracked controllers. The Vicon motion capture system is used to provide the position and pose of these controllers, which determines what areas of the jigsaw puzzle are rendered. SHIVA[^24], as mentioned earlier in §3.5.3, allows two users to interact with a large screen display using gesture, but has not been tested where “simultaneous cooperative interaction of both users” is required. In other words, just one user may point at a time, which may be OK for turn-taking games like chess, but not for true collaboration. In both of these systems, the 3D positions of the users are irrelevant.

In the special education arena, specifically in regard to those with ASDs, researchers like Cobb et al.[^29] have created “interactive contexts representing a range of social scenarios in which AS users can practice social skills.” As we noted in §2.2.2, a virtual environment may provide a secure, life-like environment in which AS users can learn and practice social skills and rules without the pitfalls of doing it in situ. In such environments as well, parameters can be controlled and adapted to vary and expand the learning experience on a case-by-case basis.

*Social Café* is one such environment, where the user is tasked with finding a place to sit in the virtual café under different conditions, such as when there are empty tables or when there are no empty tables.[^29] Other environments include Virtual Supermarket[^18], which is aimed at promoting basic shopping skills such as creating a shopping list, selecting items from the shelves, finding everything on the shopping list and paying for goods at the checkout; Virtual House[^18] for learning simple household tasks; and Virtual Transport[^18] to practice transportation-related activities.

More recently, Kandalaft et al.[^82] investigated using VR for social skills training in young adults with high-functioning autism and concluded that it is a promising tool: participants
showed significant improvement in the clinical measurements of emotion recognition and theory of mind\(^4\) as well as in real-life. The VR technology used in this study was strictly software-based, viewed on a 24-inch monitor and using a keyboard and mouse.

Other researchers have developed VE systems to help those on the spectrum overcome sensory-processing issues. Zalapa and Tentori’s SensoryPaint is one such example. SensoryPaint aims to improve body awareness of children on the spectrum, by engaging the kinesthetic sense with various activities involving a ball, like throwing it toward a moving target on the display.

While as they stand, these applications are not truly collaborative, the learning of social skills is a necessary step in that direction. Collaboration can be added as more advanced levels for those who have mastered the more basic skills, for a more encompassing educational system.

Combining the belief that in situ training is best for those with ASDs\(^6\) with their difficulty in generalizing, it may be that using an environment more approximating real-life, such as a CAVE\(^7\) may be more relevant and show greater results. The potential for specialized VR applications in a collaborative setting is vast, particularly for those on the autism spectrum.

### 5.2 Art Gallery

To get an understanding of the way the Vicon motion capture/CAVE\(^7\) system works and tracks the location of objects/people, we implemented an educational game based on the Art Gallery problem, for which we are only interested in the \(x\)-\(y\) position of each user. This is used as a stepping stone towards the second phase of the project which involves gestures.

The Art Gallery problem is a well-known computer science problem based on the real-world problem of guarding an art gallery: given a concave polygonal space\(^4\) (i.e., the

---

\(^3\)“The ability to generate inferences about the thoughts and feelings of others”\(^8\), one area of difficulty in those across the autism spectrum.

\(^4\)All convex polygons need just one guard.
where must guards be stationed so that all parts of the gallery are visible to at least one guard? The goal is to minimize the number of guards needed to protect the gallery, i.e., minimize the number of guards necessary for the entire gallery to be visible by at least one guard. Although some concave polygons can be protected with as few as a single guard, we’ve chosen to specialize our implementation for two guards/users. More than two users would require intricate shapes (see Figure 5.1) that would not fit comfortably within the allocated area.

![Figure 5.1: Art gallery shape requiring a minimum of three guards. Two guards are placed in (b) marked A and B. Both guards see the green area, only A sees the yellow area, only B sees the blue area, and neither sees the white area.](image)

### 5.2.1 Game Play

Two people, we’ll call them Jamie and Drew, are the guards for a gallery space as shown in Figure 5.2. We imagine the following interaction may transpire.

---

5 According to Chvátal’s art gallery theorem, if the museum has $n$ walls, then at most $\lfloor n/3 \rfloor$ guards are needed to supervise the gallery.\textsuperscript{[185]}

6 Think of a ‘V’ shape, which would need a single guard at the base of the ‘V’.
1. Jamie and Drew are given instructions and proceed with their mission.

2. They decide to experiment, and Drew goes to the near/bottom left corner while Jamie goes to the diagonally opposite corner.

3. “Hey! We can only see the two end rooms,” they exclaim. “We need someone to protect the center room as well.”

4. They both go into the center room, but now that is all they can see.

5. Drew says, “Let’s see what happens if you go in the corner.”

6. Jamie does so, moving toward the right (near/bottom) corner, and can now see both the center and right rooms.

7. Jamie says, “Look! We only need to protect that last room.”

8. Drew quickly realizes that he should go (anywhere) in the unprotected room, and does so.

9. They’ve successfully placed themselves to protect all three rooms, and are congratulated.

5.2.2 Technical Details

The selected polygon representing the shape of the gallery is displayed on the floor of our 12-camera/2-screen\(^7\) Vicon motion capture/CAVE\(^{TM}\) system. The wall projection is used to give instructions (Figure 5.3), as well as hints as needed (Figure 5.4).

Players wear reflective markers to make them trackable as they move about the gallery trying to place themselves to ensure security of the entire area. For simplification (and as per the theoretical definition), we pretend that our human guards can see 360 degrees about

\(^7\)One wall and the floor.
Figure 5.2: Sample art gallery floor, with walls along the edges of the light-colored shape. There are essentially three rooms that need protecting: the left room, the right room, and the connecting center room. The only way to protect this particular layout with two guards is to have one at the juncture between two rooms, and the second anywhere in the third room.

Figure 5.3: Gallery application startup screens
Figure 5.4: Being given a hint in the gallery application. The hint says, “Only the red section is visible by the guards. Place yourselves so that the white area is visible too.” Note, the arrow on the screen (pointing to the white area) represents the section of the gallery currently not seen by the guards. (The child represents one of the guards, while this researcher, standing diagonally across from the girl, represents the second guard. This configuration of guards leaves the white section unseen.)

themselves. As they move around, the area that they “protect” gets colored in red on the floor projection, while the rest remains white. When the players (guards) are situated so that the entire polygon is visible, a congratulatory message is displayed on the wall, as seen in Figure 5.5(a). In our simplified implementation, we had just one, non-customized, hint that was displayed when guards were not positioned correctly.

Setup involves the users donning a mortarboard-style hat with three reflective markers, the minimum necessary for the Vicon motion capture system to recognize as a trackable object. We do not put markers all over a person as described by Scarlatos and Friedman[144], since we do not receive position information for each marker, but rather for a grouping of at least three markers. The $x,y$ position of the marker gives the position of the person on the floor, which is all the data we need for this application.

Someone who is not participating in the application calibrates the system by selecting each grouping of markers as an object on the tracking computer. This takes about 30

8For future versions, we may explore the use of different colors to clearly demarcate each guard’s protected space.
Figure 5.5: (a) When the entire gallery is protected, players are given a congratulatory message. (Since the art gallery was merely an intermediary step for us, we did not spend time on fancy graphics here.) (b) In this close-up view of (a), the debugging mark at the position of the girl is clearly seen, albeit partially projected on the girl’s right leg.

seconds. Having the markers atop a hat helps prevent occlusion. This worked as we hoped, and occlusion was not a problem in this application.

The lower and upper bounds of the working area were indicated with additional sets of markers to help translate the values arriving from the Vicon system into user position values that are more conducive to graphics programming. This is explained in some more detail in §5.3.3.2. As seen in Figure 5.5(b), the blue dots displayed at the $x, y$ position of the users for debugging purposes are fairly accurate.

### 5.2.3 Discussion

As discussed in §2.4, the physical act of moving about within an environment allows a person to perceive things from different perspectives, and that’s exactly what we expect to happen here. As players move about, the computer helps them visualize what they would see if it were a real life scenario (by shading some parts red), thereby enabling them to analyze how the positions of themselves and others affect what areas are protected. Movement, and the
resultant change of perspective, triggers the discussion which ultimately leads to a solution.

It is clear, especially from points (3), (5), (7), and (8) in the sample dialog on page 103, that the collaboration has fostered metacognitive behaviors: in (3) and (8) they display a clear sense of understanding of the problem, and in (5) and (7) they plan a course of action which leads to a solution. These metacognitive thought processes help keep the children on track by inducing a higher level of immersion in the activity through the engagement of their intellect.

While we did not bring in groups of children to test this application, this researcher and her then-four-year-old daughter used the system and verified that it will work for users of a variety of heights. Even with a four year old’s attention span, she was able to stand still for the 30 seconds it took to calibrate the system. Being able to accommodate people of all sizes as well as calibration time are both concerns which must be taken into account when working with children.

While I was eager to move on to gesture-tracking, the art gallery application certainly has a place in education (computational geometry). To improve the application’s usability, and to enable it to better support collaborative learning, the hint module would need to be expanded. Currently, a single hint has been developed, and although applicable at all times, it’s not very specific, merely stating, “Only the red section is visible by the guards. Place yourselves so that the white area is visible, too.”

In Figure 5.6(a) not only would it be helpful if the displayed graphic reflected the state of the art gallery, but also if the guards were given additional information like, “Both guards are protecting the center area, but no guard is protecting this area. [Arrow pointing at the right room.] Discuss how it would be possible to protect that room by moving just one guard.” Likewise, for Figure 5.6(b) additional information might be, “One guard is not standing inside the gallery (white/red) area. Both guards are needed inside the gallery to protect the entire space.”

Once we had the art gallery application working, we moved onto the next phase of the
5.3 Classification Application

We chose to design and implement an application that asks users to classify a group of objects. Classification is an essential skill that is called into play throughout a child’s education, and is a component of the math and science standards.\cite{111} In fact, classifying is a necessary part of formal education through the college years. In a required core course on introductory geology, students are required to sort rocks into three categories: igneous, metamorphic and sedimentary. Art class also required sorting: impressionism, expressionism, etc. Even before a child’s formal education begins, they are already noticing that objects in the universe are not all alike, and have distinct attributes.

The foundation for categorizing objects is in place even before birth. To group items, we must first differentiate, and newborns younger than four days old have been shown to distinguish their own mother’s voice from that of strangers’ voices. There is evidence that fetuses can make this sort of distinction as well.\cite{177}

Because classifying objects is ingrained in us and plays such a vital role in education,
we’ve designed a system where students sort a group of objects by predefined attributes. Our current implementation has students classify produce as fruit or vegetables, but this can easily be modified to accommodate almost anything—shapes, animals, trees, rocks, etc.

We now look at a sample round of game play, both generally and in detail, feedback that is given during the game, and examine the technical details involved in the application. We then discuss the educational benefits of the system and potential improvements.

5.3.1 Game Play

5.3.1.1 Overview

Let’s look at a scenario where two people, we’ll call them Magenta and Goldenrod, are playing the fruit/vegetable sorting game:

1. At the start of the game, Magenta is standing on a photo of an orange and Goldenrod is standing on the blueberries.

2. Magenta (on the orange) points toward the fruit basket and Goldenrod (on blueberries) points at the vegetable basket.

3. Magenta’s orange is moved from the floor up into the fruit basket. Nothing happens to the blueberries—it remains in its place on the floor—as Goldenrod is not pointing at the correct basket.

4. Magenta suggests to Goldenrod that perhaps blueberries are not a vegetable, but a fruit.

5. Goldenrod realizes Magenta is correct, and points toward the fruit basket . . .

6. . . . while Magenta hops over to the grapes and also points to the fruit basket.

7. Both the blueberries and the grapes move up into the fruit basket.
8. They continue moving about the floor, aiding each other through discussion along the way, and sending the produce to the correct bins, until there are no more left on the floor.

9. At that point, the round of play has been successfully completed.

5.3.1.2 Details

When the application is first started, eight objects to be categorized are displayed on the floor. On the wall are displayed instructions as well as “bins” for the items to be sorted into. In our case, we display a mixture of eight fruits and vegetables on the floor. On the wall are displayed two baskets, one marked “FRUIT” and the other marked “VEGETABLES” along with the instructions, “Stand on a fruit or vegetable, then point to the basket where the item belongs.” See Figure 5.7 for a screenshot of the startup wall and floor combined—the top half is displayed on the wall, the bottom half on the floor. See Figure 5.8 for a photo of the system in use.

As users move about the floor and point at different locations on the wall, small colored circles are displayed at the position toward which they are pointing (see Figure 5.9). In the included screenshots, these markers are colored magenta and goldenrod (hence the names), with additional users necessitating additional colors.

When the application is started up, placeholders are displayed on the wall above the baskets which make it clear that four items go into each basket. Similar placeholders are displayed on the floor in place of an item once it has been moved into the correct basket for illustrative purposes. These placeholders can easily be removed.

Users continue to play until there are no more items left, as shown in Figure 5.10. If they’d like to play again, the system is restarted and another set of eight fruit and vegetables appears on the floor for them, below emptied fruit and vegetable baskets.

---

9 Screenshots are used for illustrating game play for clarity, despite not displaying user locations.
10 No means to ask for a new game has been implemented, but this is a rather simple modification.
Figure 5.7: Classification application start screen

Figure 5.8: An adult and a child working together to classify the fruits and vegetables displayed.
5.3.2 Feedback

In any educational activity, on the computer or off, a student needs feedback to learn, and we have incorporated feedback in a few different ways. When you talk to someone, you know they’ve heard you because they respond to you. With a computer, you want to know that the computer has “heard” you and registered your request. For that, we display a dot on the wall toward which the user is pointing, as mentioned earlier in the game play section (§5.3.1.2). These dots are displayed using dissimilar colors for each dot so as to avoid confusion.\footnote{Unfortunately, we did not refer to any guidelines for choosing colors to avoid ambiguity for the color-blind.}

Although not implemented in their current system, Störring et al. mention displaying a 3D line—visible with stereo-glasses—from the user’s pointing finger to the object the
computer believes they are pointing toward.\cite{165} The user then knows how to correct their gesture. Although this would make disparate colors unnecessary, their system has been built for individual use. It is unclear if—and at what cost—such an approach can be extended for multiple users. For these and other reasons, we have chosen not to use special glasses, but believe that since the dot is at the point at which each user is pointing, it should be clear which mark belongs to which player. Informal testing has not shown this to be a problem.

There may be ambiguity when two players are pointing at the same—or close—position, but there is only minimal chance that this will create confusion, e.g., if the pointers are near each other, the users are probably pointing at the same thing. Another area for possible confusion exists when one or more users is standing too near the screen and can’t easily see their pointing trajectory. Our approach to preventing this problem is discussed in \S5.3.3.2.
Providing feedback on the floor in close proximity to each user, while unambiguous, interrupts focus and was therefore never considered.

The dots can be a source of other kinds of feedback as well; one variation we tried displayed a check on the dot if pointing at the correct bin, and an X otherwise. This can prevent users from wondering why their item is not being placed into the bin—“Is the system not working or am I wrong?” Ultimately we took this off as we want users to pay attention to the item they are placing and the bins: the primary focus should not be on the dots. The quiver of the dot from a person’s naturally imperfect hand steadiness is usually enough to let the user know that the machine is, in fact, responding.

A fruit-vegetable sorter is trivial as an application: either you know it, or you don’t but guess, and with practice will remember it. With other classifications, such as rocks, the students are expected to recognize specific attributes of the items they are looking at. Because of the various complications involved in implementing a help function, feedback of this sort is left for further research.

Although the plan was to add an audio component to give feedback, such as stating “that is not a vegetable” when a user is standing on a fruit yet pointing at the vegetable basket, we have discovered various difficulties in implementing such a component. The decision to leave audio out of our system was made after careful consideration. First of all, being told you’re making an error every time you make a mistake can become discouraging. Although it is sometimes helpful to know that you are doing something wrong before you complete a task, real-time error-checking already provides that functionality by not allowing placement into a wrong bin. Adding positive audio feedback (“that is a fruit, good job!”) to offset any negative feedback would create a ruckus. Constant chatter from the computer is distracting for even those with good focus; how much worse it would be for someone who is highly distractible.

Furthermore, it is unclear to whom the audio feedback is intended. If one person is standing on cauliflower and the other on grapes, yet both are pointing at the fruit basket,
the computer would be saying, “that is a fruit, good job! ... that is not a fruit” or vice versa. One way to approach this issue may be to use directed comments: “Magenta, that is a fruit, good job! ... Goldenrod, that is not a fruit,” or “cauliflower is not a fruit ... grapes are a fruit, good job!”

However, this makes the computer even chattier, which prolongs the distraction for those with attention difficulties, even as it becomes less of a guide for others, as running commentary fades more easily into the background than does short, directed help.

Additionally, none of the three approaches to handling multiple queued audio streams seemed well-suited to our application:

- Cutting off the current stream in place of a newer one is messy, and an important message may be missed.

- Queueing new streams may miss streaming audio while it is relevant, and may confuse users when it does play.

- Throwing out new streams while another stream is playing may mean users miss important feedback.

One idea we are looking into is streaming personalized feedback to Bluetooth headphones. This would ensure that each participant receives congratulatory—and error—messages meant only for them, and would minimize general audio feedback. Although the handling of audio streams may still be a problem, especially among fast-moving participants, the minimized audio reduces the problem enormously. As more people join together to collaborate, Bluetooth headphones become even more necessary.

By removing distracting feedback, Bluetooth headphones may enhance collaboration in a number of ways: minimizing user-specific feedback\textsuperscript{12}, providing an environment where

\textsuperscript{12}The amount of general user feedback (i.e., feedback sent to the speakers for all to hear) may be seen as a constraint in a collaborative environment. Minimizing it allows more people to participate.
discussion can take place without being overspoken by the “guide-on-the-side,” and keeping negative feedback from being known publicly.\textsuperscript{13}

However, we do anticipate a couple of concerns that may arise. There is the possibility that the headphones will interfere with collaboration—not being privy to another’s feedback is, by definition, exclusionary. Putting up this wall may create others, possibly causing participants to focus on their individualized feedback and tune each other out. On the other hand, Bluetooth headphones may directly support collaboration with personalized feedback like, “Your partner seems to be having some trouble. Are there any suggestions you can make to help them move along?”

A further concern with Bluetooth headphones is one we have mentioned in the previous section: the more equipment that’s required, the less feasible the system may be for those with various disorders.

Unfortunately, as with most computer-based systems, some types of feedback can not be provided. Because all objects should be displayed as roughly the same size, one where the object can be easily seen from a few feet away, neither absolute nor relative size can be ascertained. Texture, too, which is often used as a differentiating factor, is lost. To counteract this problem, actual specimens may be placed near or within the environment, where feasible. Aside from allowing participants to get a better sense of the objects, the objects’ presence is expected to facilitate discussion.

\subsection*{5.3.3 Technical Details}

The technical aspects of the system can be broken down into tracking/capture, display, and process, which we discuss in this section, together with known technical problems, some which we have addressed, and others which need to be addressed.

\footnote{\textsuperscript{13}As we mentioned in the section on collaborative learning (\textsection 2.3.1), people within a collaborative group should be at about the same level. However, it may be impossible to control this in various settings. Using Bluetooth headphones to give feedback will keep negative feedback from being obvious to the other collaborator(s), preventing them from treating lower-performing peers in a condescending manner.}
5.3.3.1 Capture

Our system is comprised of a front-projected floor and a single rear-projected wall CAVE\textsuperscript{TM}, surrounded by 12 Vicon cameras, three of which are visible at the top of Figure 5.11. These specialized cameras project infrared light which is reflected off the markers worn by the subjects, and filter out all light except infrared. To aid in reflector detection, ambient light must be kept to a minimum. Accordingly, walls are painted black, and dark navy or black clothing should be worn in the environment.

The Vicon motion capture system receives input from the cameras and passes the input along. The TrackD server grabs this data and relays it to another computer where it is placed into shared memory. We use a Max/MSP patch (Figure 5.12) written by a third-party to obtain the data and pass it along to the application, where it is processed in real time. Max/MSP is a graphical programming environment used by musicians and artists. Since the Vicon/CAVE was used primarily by art majors, we were able to have their patch customized for us.
The points we are tracking, each labeled with a set of three reflective markers, include two points per person and two calibration points. The two points we track on each person is their shoulder and wrist of their dominant hand (or the hand they’ll use for pointing). In our informal tests of three people of heights ranging from under four feet to almost six feet, tracking the shoulder position was more accurate than tracking a point on top of the head. Our “feedback dot” appeared to the user to be displayed at the precise location they were pointing toward. Placing the reflective markers on the shoulder and wrist are less likely than the hat to be a distraction for those with ASD-related sensitivities, since the hat is additional attire, while sensors can be placed directly onto the child’s own clothing at the shoulder and wrist positions. Two calibration points are used, one at the front bottom left of the environment, the other at the rear top right. The points we are tracking can be seen in Figure 5.11. Each group of three markers is selected as an object within the Vicon motion capture application, as seen in Figure 5.13.
5.3.3.2 Display and Process

From a group of fruit and vegetables, we randomly select four of each of them, and randomly display them on the floor. Although the proportion does not necessarily have to be 50-50, we chose to display equal numbers of each since they fit nicely on the wall, when categorized: four in the basket on the left, four in the basket on the right. Although we chose to give the learners eight objects to sort for this demonstration, up to fifteen items could be comfortably placed on the floor. These items are spaced far enough apart to decrease the likelihood that the capture system will be confused, and large enough so that even if the user forgets what they are standing on, a quick glance down is all that's necessary. In most cases, it should not be necessary to move to the side to get a complete visual. Items are displayed in the far two-thirds of the screen, leaving one-third of the floor blank between the players and the wall display, to prevent interference with the natural pointing gesture. Refer back to Figure 5.8 for relative size and placement.

The wall displays two baskets, but these may be substituted with garages, toy chests, or animal pens, to conform to the objects being classified. Two bins is not an absolute
requirement, but we recommend no more than six, in a three across by two down grid, to avoid accuracy issues.

While any number of people that will reasonably fit in the floor area can be set up with markers for game play, two or three seems to be optimal within the proportions of our CAVETM. This allows each person enough space to move around independently without bumping into anyone else, while also giving them the opportunity to sort a moderate number of items. Furthermore, discussions are more focused when they’re one-on-one or in a small group. At the other end of the spectrum, individuals can use the system as well, but working by oneself, one misses the non-trivial benefit of collaboration.

As in the Art Gallery application, we use two additional sets of markers as calibration points to indicate the lower and upper bounds of the working area, as seen in Figure 5.14. We considered marker ‘B’ at the front bottom left, to be the lower bound at position (0, 0, 0), and marker ‘C’ at the rear top right, to be the upper bound at position (800, 600, 400). Using Algorithm 2, we normalized the data to be within this range. Any data outside the boundaries was discarded. Although rigorous testing to ensure accuracy was not done, the resulting data worked as expected.

For every set of user data that arrives, the intersection of the line defined by the two

---

14 Are eight people really necessary to sort eight items?
Algorithm 2 Processing position data.

```plaintext
if corners not computed then
    for each $i$ in $x, y, z$ do
        $d[i] \leftarrow upperbound[i] - lowerbound[i]$
    end for
end if
for each person do
    for each personmarker do
        for each $i$ in $x, y, z$ do
            $pos[person][personmarker][i] \leftarrow (viconvalue[i] - lowerbound[i]) \times (MAX[i]/d[i])$
        end for
    end for
end for
```

points on the arm and the wall plane is computed. The point of intersection is where that user is pointing, and a large dot is displayed, as discussed. If the line does not intersect the plane, the user is not pointing at the wall screen, and the data is ignored. If the line does intersect the plane, they may still not be pointing at the screen, but rather at some point to the side, above it, or the floor, so we ignore any data outside the constraints of the screen. We then determine toward which section of the wall they are pointing, and if it’s correct for the item they’re standing on (as determined by the location of the shoulder), the item is moved into the bin.

We do not require the user to point at a bin for a specific length of time as in other applications (as mentioned by Carbini et al.[24]), as we are not waiting for a “click” but rather a “point”. This has its drawbacks, as discussed in the next section.

5.3.3.3 Known Problems

All vision-based systems are going to be subject to some degree of occlusion. Although increasing the number of markers per object helps insure against this, it can “result in exponentially increasing the ‘confusion factor’, i.e., keeping track of which marker is which.”[172] The Vicon system, suffering from occlusion-related confusion, repeatedly confused objects (a hand for a shoulder, one person’s shoulder for another’s), even when not
in close proximity to each other, but especially when close by. Because this often occurred during calibration, we adjusted our code to use a string of values, representing the correct ordering of objects, to override what the Vicon system seemed to believe. Yet when this confusion occurs during game play, we need to restart the system and re-mark the points to be tracked. This has a tendency to happen more with smaller kids, as their limbs are short, so markers are closer together.

Bartoli et al.’s design guidelines for ASD children recommend minimizing transition time or risking a child losing concentration[10], so certainly the loss of object mapping during game play would be an issue.

We discovered an error that occurs somewhere between the Vicon system and the TrackD server which acquires the camera data and makes it available to end-user applications. Translations applied to one object affected both that object as well as the first object in the stream. Because we do not have access to the program code, we were forced to implement a workaround. We resort to tracking a dummy point (labeled ‘A’ in Figures 5.14(a) and 5.14(b)) as the first object in the stream, and then discard this data.

While the wall display was rear-projected, the floor used an overhead projector, and was therefore subject to obscuration, as the users are physically on top of the display surface. For each user, the obscured part would include the section of the floor that they were standing on, plus the shadow of their body and outstretched arm. Narrow arm-shaped shadows were not typically a problem and, as people don’t generally lock their elbows when pointing at something close by, the shadow was shorter than arm’s length (see Figure 5.15(a)). We tried to make the item images large enough to be seen around body shadows, although in general, the centered ceiling projector created shadows outward, which were non-blocking, as seen in Figure 5.15(b). Additionally, people standing in positions which otherwise might cast long shadows on those behind—front row center—had significantly shorter shadows because the light source was projecting straight down on them.

Using gestures, while natural to humans, can be physically draining if required over long
Figure 5.15: Shadow effects. (a) Cut-out from Figure 5.8: arrow pointing to shadow of the child’s arm, which is entirely contained within the bounding box of the item that was just placed into a bin. (b) Although outstretched toward the display, my arm’s shadow is barely visible on the floor.

periods of time\textsuperscript{15,20}, especially if they have to be exaggerated for recognition by a computer. For example, our system requires that we move our hands farther out and maintain their lift, in order to distinguish the deliberate gestures from casual ones, which typically aim toward the floor.

In our informal tests, we have not noticed an incident where the user’s hand crossed

\textsuperscript{15}Cabral et al.,\textsuperscript{20} concedes however that “it is convenient for sporadic use in virtual environments.”
over the wrong basket to get to the correct one. This may be because our natural pose is not to be pointing, and we put our hands down while walking about. When raising the hand to point, it is usually in the predetermined pointing direction. However, since we only recognize a pointing gesture, it is very easy to just move one’s hand around, pointing at different locations on the screen and the item will move into the correct bin at the moment that the pointing is correct. This makes it possible to place all objects in the correct bins without learning, or knowing, anything, if done purposefully. A possible solution is discussed in §6.3.

5.3.4 Discussion

There are many advantages to this style of learning. Learners do not need to wait for their turn. They move about the workspace, stimulating their kinesthetic sense, thereby engaging more of their body in the process. They work at their own pace, sorting whatever objects they are familiar with, while freely discussing their task with other players or onlookers. Although one player can conceivably classify all of the items while the others watch from the sidelines, we do not include score-keeping. This is designed to encourage collaboration among the players to work toward the common goal of “clearing the floor”, and not be concerned with how many “points” they accumulate. (See more below.)

An enforced collaboration option may be useful in some populations, such as among those with ASDs, to require participants to work together, as discussed in §2.3.2. This may work in a couple of different ways: splitting the floor into two identical workspaces of half the number of items to sort, and requiring both players to stand on the same object and classify it together (requiring them to work in concert with one another), or not allowing the game to proceed until both players have classified an object (encouraging players to assist one another). It may be advantageous to provide both options.

Whereas collaboration among children sorting fruit and vegetables might seem unnecessary, especially with feedback indicating whether the placement is right or wrong,
advanced topics, like sorting rocks for an introductory geology class, are more likely to trigger discussions on what to look for. A rock classifying game may have students sorting rocks into igneous, metamorphic, and sedimentary bins. A sample game with Violet and Hunter may go like this:

1. Violet stands on a photo of sandstone, and Hunter on gneiss.

2. Violet points to place the sandstone into the sedimentary bin, while Hunter points at the igneous bin.

3. The sandstone goes into the sedimentary bin, but the gneiss remains on the floor.

4. Violet says, “Notice the layers in the rock. I think that’s called foliation, which is an indication that the rock has metamorphosed.”

5. Hunter responds, “I see! It must be metamorphic,” and points to place the gneiss into the correct bin.

In this sample interaction, we see how the two players are collaborating, helping to point out the various factors that go into making a determination on an item’s classification. Metacognitive events show up as analyzing in point (4) and understanding in point (5). Increased immersion in the activity brought about by engaging the child’s intellect—by stimulating metacognitive thought processes through collaboration, combined with arousing the child’s kinesthetic sense, creates what we believe is a synergistic effect on learning.

Since participants in our system are co-located, they can talk naturally without being mindful of the requirements imposed by communicating in real-time across a network, such as facing a camera or taking extra care to speak loudly and clearly for the microphone, that is necessary in the NICE project[135] and others that are networked. Overall, we have tried to keep the focus of the users off the technology, and on task, by avoiding distracting stimuli or requirements.
Although we have not included a score-keeping feature, this may be an interesting game option to add. Keeping score can encourage collaboration if done across multiple rounds of game play (first one team and then another use the system), particularly if the awarding of points is tied to collaboration-related metacognitive events.

Another improvement would remove the restriction on the quantity of objects that may be classified. Rather than limiting quantity to what can be comfortably displayed on the floor, new objects appear on the floor once a user has correctly placed an object and moved away. One possible application using this feature may be in Geography: classifying countries by continent. Displaying a world map on the wall, with the continent’s location acting as its bin, enables users to sort all the countries of the world (or some subset thereof), as desired by the teacher. To begin, the original eight or so countries are displayed on the floor. Then, as each correct classification is made, the floor is updated from a pool of countries until the pool is depleted, at which point the game is over. One consideration would be how to display all classified items “in” the bins on the wall.

Competitive score-keeping within a round makes more sense in an unlimited-quantity classification game than it does in a game with all objects pre-displayed. For those on the ASD spectrum, competitive score-keeping may be a predecessor to working in a more collaborative fashion. Score-keeping may provide the motivation necessary to maintain interest while the ASD user becomes familiar with the system, an essential step if they are to successfully collaborate in future levels.

As with the Art Gallery, as of yet, we have not brought in groups of people to test the system and its impact on learning. Roussou[138] discusses the difficulty in evaluating VR applications, both in a museum setting (diverting them from planned activities) as well as a VR laboratory setting (logistics of getting users to the lab), and Piper et al.[124] discusses hurdles specific to testing special needs students (multiple levels of approval were necessary, took months to build rapport). Means of overcoming these hurdles need to be explored so

\[16\] Perhaps collaborating with a special education school on a long-term project.
that we can proceed with some interesting experiments we plan to conduct, as discussed in §6.3.
Chapter 6. Conclusion

This dissertation explores ways to support collaborative learning in two types of computer-enhanced environments, tangible user interfaces and virtual reality. Supporting collaborative learning within these active environments contributes not only to the child’s academic development, but their social development as well.

6.1 Contributions

The goal of these projects has been to create environments that would allow children to be immersed in and engaged with the learning process in a collaborative manner. To that end, we have partially succeeded. Working with Scarlatos to build and install the TICLE tangram, we showed that such an approach can help keep kids focused on a learning goal, while enjoying the activity. The computer in the experimental group, acting as a guide-on-the-side, provided the right amount of support for keeping children engaged (even as children in the control group began to get restless and lose interest), yet not so much that it became the center of attention. Although at times one child would decide that she wants to “try something,” for the most part, children worked together and discussed how they ought to proceed.

Problem solving, in part, requires that people learn that (in all aspects of life) the first approach isn’t necessarily the correct approach. Despite the fact that the Art Gallery application was intended merely as a learning phase, it provides students with the opportunity to explore how a change in perspective may contribute toward finding a solution to a problem: moving about the gallery, analyzing each perspective, and comparing
it with other perspectives. The rudimentary guide-on-the-side supports learning by giving participants food for thought, encouraging metacognitive discussions. Presenting the task as a collaborative one that requires physically active involvement compels all participants to do their share. The resulting full-body engagement in the learning process, coupled with the support of both collaborator and guide-on-the-side, is the motivation and means to learn.

Although the CAVE/Vicon classification system has not been tested on students, we have shown that it is capable of tracking and responding to the gestures of multiple children in a virtual reality learning environment. This enables the creation of learning activities that are truly active and collaborative. In comparison with socio-ec(h)o\[182\, 183\], which is most similar to this work: while they track a single group of five markers to determine each person’s position \((x, y)\) and posture—standing or crouching \((z)\), we track the \(x, y, z\) coordinates of two distinct parts of each participant’s body to make sense of their gestures as they move freely about the environment. Feedback is based on position, orientation, and gestures to help affect a solution to the task.

We have put together a number of separate systems (immersive environment, gesture recognition, guide-on-the-side) to form a cohesive system ideal for use in collaborative educational settings, with particular potential for special education. We have also discussed the various considerations that must be taken into account when developing such systems for general, as well as special education, and presented solutions to some of them. Additional considerations are discussed in this chapter.

### 6.2 Limitations, Challenges, and Open Issues

The challenges we encountered while working on our three-dimensional tangible user interface were discussed in Chapter 4. Here we focus on the Vicon/CAVE application.

While the exercise is engaging, the flaws in the technology prevent users from being fully
immersed. The inaccuracy of the Vicon motion capture system created confusion and caused frustration at times. Although it didn't happen often, two or more of the points we were tracking sometimes got confused, and we would need to restart the application. Because these remaining technical issues can detract from the experience, the calibration process must be streamlined, and the response time and accuracy of the system must be improved.

The attire requirements of cap and wristband could be an issue within certain populations, such as among those with an ASD, who may resist certain sensory inputs, including wearing garments of specific colors or textures. Although this concern can be dealt with, it may require considerable time for them to acclimate to unfamiliar sensations, however unnoticeable to the general population. Adding technology, like individualized feedback via Bluetooth headphones or stereoscopic glasses, may exacerbate sensitivities and interfere with collaboration.

Although it may feel like a glaring omission and detracts from the users’ immersion in the environment, for all the reasons noted above in a previous chapter, we have left out all audio, and leave its inclusion for future work. In the application we have implemented, this omission is tolerable—feedback is given on the display, towards which users face at all times (they are pointing at it). However, in applications that rely more on movement, this absence may be more evident.

Also as we have mentioned earlier, it is impossible not to obscure at least part of what a user is standing on, although we try to minimize the impact of the obscuration by displaying images as large as reasonably possible[^1]

The size of the floor display is also a limitation, allowing a maximum of two to four people to comfortably move around in, depending on the activity; collaborating on an active movement exercise may require more space. This is a Vicon and CAVE limitation and can only be addressed by more advanced hardware.

[^1]: We need to maintain scale to some degree—fruit like blueberries would be unrecognizable if displayed too large.
6.3 Future Work

There are a number of directions in which we can proceed with our research, and here we take a quick look at some of them.

- **Audio.** As we discussed in an earlier chapter (§5.3.2), we declined to include an audio component. We do, however, find this to be a glaring omission, and plan to research various alternate approaches, such as streaming personalized feedback to Bluetooth headphones to ensure participants receive congratulatory and error messages meant only for them.

- **Guide-on-the-Side.** A guide-on-the-side to promote collaboration and learning would be an important addition to our system, and needs to be investigated further to find an ideal balance of helpfulness, while minimizing any distraction. Some aspects of a guide-on-the-side that could be explored include:
  
  - **Speech/voice recognition.** Recognizing spoken commands will make the system available to more users, particularly those with physical handicaps, and differentiating among users’ voices would allow targeted feedback.
  
  - **Facial expression recognition.** Recognizing when a user is stressed, anxious, or confused, can allow the guide-on-the-side to respond more appropriately.
  
  - **Instructors’ interface.** Discussed by Scarlato and Scarlato [151], an instructors’ interface would allow teachers to customize the application to support a broader age group and range of sorting activities.

- **Gesturing.** Recognizing only the pointing gesture leaves the door open to “cheating”, i.e., simply moving one’s hand about, pointing willy nilly at the bins until it’s the correct bin. Requiring some version of a “click” would prevent this. Recognizing more complex gestures—both static and dynamic, not restricting pointing to one
plane, and understanding gestures when users interact with physical objects within the environment, would make for a more enriching experience, and should be investigated.

- **Increasing Accessibility.** Cost and space considerations make the system unavailable to many learners. To maximize the number of learners that have access to this technology and can reap its benefits, these considerations would have to be addressed. Microsoft’s Kinect, discussed in §3.5.3 is significantly more affordable than the Vicon motion capture system and, although it has drawbacks which may make it unusable in such a system, investigating it may yield some interesting results. Looking into ways to shrink the system without compromising on immersion, movement, and collaboration, will also contribute to making the system more accessible.

- **Experimentation.** Bringing in small groups of students, ranging from kindergarten through college, to measure impact on learning, is work we plan to do as well. The experiments we plan to conduct tie in to the themes of Chapter 2: Education, Social Interaction and Collaboration, and Movement.

  - **Education.** As we have done with the TICLE system, performing an analysis of cognitive and metacognitive events will give us an idea of the level of learning that takes place and will enable us to provide the kind of assistance that is needed. Also, pre- and post- tests should be done to measure learning, transfer of knowledge, and ability to strategize.

  - **Social Interaction.** Some questions we hope to answer by customizing our system: can the scaffolding contribute to improvement in social interaction for ASD children with regular usage? Can social norms (e.g., personal space) be taught? Can the system teach ASD children to understand or ignore the social signals of neurotypical (i.e., non-ASD) children so they can operate in a shared environment without feeling threatened? Can computer-supported collaboration provide shy children the positive social interaction experiences needed to promote
better interaction in other social situations?

– **Collaboration.** Does the system have more of an impact on one gender than the other? Does using the experimental system enable users to collaborate on a higher level? Assess the quality of collaboration: is one teammate impatient with the other? Does a teammate react poorly to the other’s comments? Is a more knowledgeable teammate withholding information they should be sharing? Can these issues be tackled with improved hints and incentives? How can SSRL be implemented? Can a speech recognition module recognize and respond to negative collaboration tactics to improve collaboration and increase learning? Also, how is collaboration impacted when adding a third user?

– **Movement.** Compare and contrast the two types of movement we have explored (manipulating objects with the hands vs. manipulating environment with the whole body) and assess their impact on learning in various groups (neurotypical children, ASD children, hyperactive children, boys, girls, etc.).

One of our immersive applications relied on the change of perspective to solve the task, while the other exploited user gestures while moving about. An application which combines the two, requiring gesturing while also analyzing perspective, we believe, would make full use of the body’s kinesthetic abilities. It would also be interesting to attempt to combine the two approaches: using tangible objects within an immersive environment, but we have yet to find a compelling application.
Appendix A. Solving Instant Insanity

Figure A.1: Instant Insanity cubes.

Begin with a set of instant insanity cubes, as in Figure A.1. The object of the game is to put the cubes together such that each side of the solution contains one of each of the four colors present (Figure A.2).

Figure A.2: Instant Insanity solved. We could see that both displayed faces—the top and front—have one cube of each color.
This can be solved by trial-and-error or by drawing some graphs, with the graph approach having the advantage of making known if there is a solution before we even begin. Start by creating four graphs—one for each cube—that contain four vertices, each representing one of the four colors. Since the cubes have six faces each, there are three pairs of opposite faces, and we represent these relationships with edges on the graph. For example, on my first piece (arbitrarily chosen), there is a red side opposite a green side, a green side opposite a yellow side, and a purple side opposite a purple side. So, as shown in Figure A.3(a), an edge connects red with green, green with yellow, and purple with purple.

Once we construct the four individual graphs as shown in Figure A.3, we combine all of the edges into one graph, as seen in Figure A.4. This graph contains all pairs of opposite faces that exist among the four cubes.

We then need to find two separate subgraphs within this graph, each containing four edges—one edge (opposite face pair) from each cube. One subgraph (set of four edges) will represent the top and bottom of the solution, while the other subgraph will represent the front and back of the solution. The remaining edge of each cube is not part of the solution—this edge represents the faces that are on the sides, touching the adjacent cubes. Additionally, each of the four vertices in each subgraph must have two edges coming out of it, for a total of four times that each color appears in the solution (once on each side). If there is no way to divide the graph of Figure A.4 into two subgraphs which fit these criteria, there is no way to solve the puzzle with that particular color arrangement. In this case, the
Figure A.4: Graph of all four instant insanity cubes combined. The numbers indicate which cube the edge came from.

(a) Front-back pairs
(b) Top-bottom pairs

Figure A.5: The graph of Figure A.4 split into two solution subgraphs. 

We now know which pairs of opposite faces work together to form the solution. For example, as seen in Figure A.5(a), cube 1 has a pair of opposite sides whose colors are red.

\[\text{If you'd like to try to break the graph of Figure A.4 into subgraphs on your own, you can try by using the edge connecting purple to purple (from cube 1), and you'll see it can not be done. Therefore, you'll know that each subgraph contains one of the other edges from cube 1. Take it from there.}\]
and green; one of these goes in front, while the other is in back. At the same time, cube 2’s yellow-green pairing has one in front and one in back, cube 3’s yellow-purple pairing has one in front and one in back, and cube 4’s red-purple pairing has one in front and one in back. This works the same for top-bottom pairings as shown in Figure A.5(b). It is then a matter of rotating the pieces 180° around any necessary axis until a solution is found.

This description is based on Ivars Peterson’s explanation [121].
Appendix B. Strings for Representing Puzzle State and Solution

To satisfy our need to assess the state of the TICLE tangram, Scarlatos designed the following means to represent polygonal puzzle pieces. This approach is not restricted to the tangram, and works with all two-dimensional polygons.

For puzzles whose pieces are polygons, assign an identifier (we’ll use a number) to each unique piece shape. Identical pieces, assuming they are interchangeable, have the same identifier. Furthermore, each edge is labeled (we’ll use letters for this), with symmetrical pieces reusing labels for equivalent sides. For example, squares have four sides, all of which are equivalent; it does not matter which way a square is rotated. Likewise for equilateral triangles and any convex equilateral polygons. If the various rotations of such a polygon must be differentiated between, each edge would require a unique label. This may occur, for example, if the pieces were colored, and the color somehow factored into the puzzle. A square with the top blue and the bottom green may be different than the top green and the bottom blue.

Some polygons, like parallelograms and rhombi, which have multiple rotations that are congruent, may have multiple (typically, but not always, opposite) edges labeled alike. Most other polygons, including isosceles triangles, have rotations which are all distinct, and therefore require separate labels for each edge.

Any two polygonal puzzle pieces must touch in one of 16 ways. Suppose edge $a$, which is line segment $[a_1, a_2]$ of piece 1 touches edge $b$ (line segment $[b_1, b_2]$) of piece 2, then for $1$ Where $a_1$ and $a_2$ are the endpoints of that line.
each of $a_1, a_2, b_1, b_2$, that endpoint either touches the other edge or does not. If it doesn’t, the value is 0, and if it does, the value is 1. These four 0/1 values are taken as binary digits that, when used together, form the adjacency representation value. For example, if the pieces are disjoint, each of $a_1, a_2, b_1, b_2$ would be 0, and 0000$_2$ is equal to 0$_{10}$. If endpoint $b_2$ is touching edge $[a_1, a_2]$, endpoints $a_1, a_2, b_1$ are not an actual part of the adjacency. This makes the adjacency representation value 0001$_2$, or 1$_{10}$. This goes up until 1111$_2$ (15$_{10}$) which represents edges of the same length touching each other completely ($a_1, a_2, b_1, b_2$ are all part of the adjacency).

Now we can construct a string of adjacencies representing the state of a puzzle. If piece 1’s side $c$ is touching piece 2’s side $b$ with adjacency value 11, it would be written as “1c.2b.11”. If piece 3’s side $c$ also touches piece 1’s side $c$, with adjacency value 7, that would be written as “1c.3c.7”. These two strings would be put together with a semicolon in between, forming “1c.2b.11;1c.3c.7” which represents the current state of the puzzle. Each component part of the string we now refer to as a substring, i.e., each substring represents a single adjacency.

In advance, we construct a representation of what the solution should look like, such as “1c.2b.11;1c.3c.7;2a.3a.15;2b.3c.15;3b.3c.9”. Then, every time the state of the puzzle is computed, we can compare each puzzle state substring with the solution string to see which adjacencies in the puzzle state occur in the puzzle solution. If the puzzle state string is identical to the solution string, the puzzle has been completed.

If some substrings of the puzzle state string appear in the solution string, there are some correct adjacencies, i.e., a partial solution. Whether a hint is given depends on how this current state differs from previous states and how long the puzzle remained in the particular state. If the adjacency had not been there before, congratulations on making progress may be presented, possibly with a next-step hint. If the state has not changed for a while, encouragement would be given, but it’s possible that the users have simply walked away. If the puzzle state seems to indicate that the users are moving away from a solution, it may be presented.

\footnote{For our purposes, disjoint pieces are ignored.}
be due to user experimentation, and refraining from giving feedback may be best. This last situation may also be indicative of confusion, and it may be best for these users to approach the problem from a new angle.

This description is taken from the Scarlatos paper, *Puzzle Piece Topology: Detecting Arrangements in Smart Objects Interfaces*\textsuperscript{[149]}, with additional detail from her paper, *TICLE: Using Multimedia Multimodal Guidance to Enhance Learning*\textsuperscript{[150]}.

This approach may be extrapolated for three dimensions as described by Scarlatos et al.\textsuperscript{[148]}
Appendix C. Interview Questions to Gauge the Effectiveness of Tangible User Interfaces as Mathematics Learning Tools

1. Have you seen this type of puzzle before? If yes, was it in the classroom, in a museum, on a test, or elsewhere? Were you successful then?

2. What do you think someone could learn from playing with the puzzle? Do you think you learned anything from today’s activities? If yes, what did you learn?

3. If you were asked to help someone else to solve this puzzle, what hints or suggestions would you give them?

4. Were your teammates friends, or acquaintances? Have you worked with them on group projects before? In the past, what sort of projects have you done successfully in a group?

5. Do you think you were given enough time to solve the puzzle, or too much time? Did you understand what you were supposed to do?

6. Was the computer interface helpful or distracting? How easy was it to work with? Were you able to ignore it when you wanted to?
7. Did you look at any of the hints? Were the hints helpful or confusing? Did you understand the questions they asked? Did you try to solve the problems they posed, or did you go straight to the answer?

8. If you could redesign the computer interface, how would you improve it? Are there other things about the puzzle or the environment that we could make better?
Bibliography


Autobiographical Statement

Shalva S. Landy has been interested in computers since the fourth grade, when she was first introduced to programming using a version of Basic. Taking a required Core course during her first semester of college reawakened that interest. As a senior, she began working with Professor Scarlatos on research described in this document. Her research interests include tracking objects and people in collaborative educational settings.

In her spare time, she enjoys reading, drawing, logic puzzles, and activities that require orderly thinking, such as sewing without a pattern.