The Equipped Explorer: Virtual Reality as a Medium for Learning

by

Scott Wilkins Greenwald

B.A., Northwestern University (2005)
M.S., Free University of Berlin (2008)
S.M., Massachusetts Institute of Technology (2010)

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning,
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2018

Signature redacted

Author . . . .

Program in Media Arts and Sciences
October 18, 2017

Signature redacted

Certified by . . . .
Pattie Maes
Professor of Media Arts and Sciences
Thesis Supervisor

Signature redacted

Accepted by . . . .
Pattie Maes
Academic Head, Program in Media Arts and Sciences
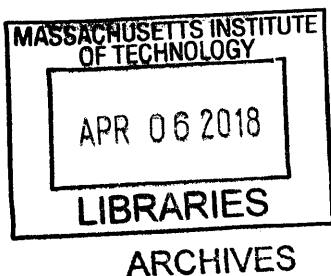
The Equipped Explorer: Virtual Reality as a Medium for Learning

by

Scott Wilkins Greenwald

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning,
on October 18, 2017, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

What opportunities does virtual reality offer to improve the way we learn? In this thesis, I investigate the ways that constructivist approaches, in particular exploratory and experiential learning, can be uniquely supported by immersive virtual worlds. Against the background of these learning theories, I introduce a design framework that centers around defining a medium of virtuality that is fundamentally social, and uses capture of movement and interaction as a key means for creating interactive scenarios and narrative. Within the world conjured by this medium, the Equipped Explorer learns, reviews, creates and communicates using tools that I propose and classify according to a taxonomy. A series of prototypes and design explorations are used as proofs of concept for aspects of the design framework. Experimental studies are used to investigate foundational questions concerning the learning benefits of using VR over 2D interactive media, and the viability of social interaction and collaboration in VR. I reflect on the implications of this framework and my experimental results to extrapolate how they might impact the future classroom and the practice of learning and discovery more broadly. Finally, I discuss what kinds of research might be needed to maximize that impact moving forward.

Thesis Supervisor: Pattie Maes
Title: Professor of Media Arts and Sciences

This doctoral thesis has been examined by a Committee of the Program in Media Arts and Sciences:

Signature redacted

Christopher M. Schmandt....................

Member, Thesis Committee
Principal Research Scientist

Signature redacted

Professor Albrecht Schmidt.................

Member, Thesis Committee
Professor of Human-Computer Interaction and Cognitive Systems
University of Stuttgart, Germany

Acknowledgments

7

Contents

List of Tables

List of Figures

Chapter 1

Introduction

As we hurtle forth into the 21st century, we must continually reevaluate what to teach and how to teach it. Shifting from a focus on information recall to information search, the older learning approaches of memorization and practicing performance have been replaced by a focus on critical thinking, creativity, and team work. Correspondingly we have shifted from "frontal instruction" to active and participatory learning, incorporating "flipped classrooms" that focus on peer interaction and team work during class time. At the same time, we have been battling problems of motivation and attention span. Students report being unmotivated to learn material that seems irrelevant to their career plans, and have trouble mustering the patience and discipline to read long texts or produce in-depth essays.

Modern instructional approaches address these problems in part: experiential learning derives an advantage from representing the "real world" in ways that make the relevance and applicability of the knowledge they convey self-evident [Kol84]. Exploratory learning seeks to harness students' curiosity to keep them focused, and simultaneously train skills of critical thinking and inquiry [Rie96].

Virtual reality offers an alternative medium in which to implement such strategies, and in this dissertation I investigate its strengths and weaknesses in reference to this application area. I hope that learners worldwide will be able to benefit from the use of virtual reality, and I consider this most likely if our collective efforts are focused on the areas where the advantages are greatest. Some

19

of these advantages are non-obvious, and others have been posited but remained unproven. This is where I view my contributions to be most valuable.

Virtual reality technology has existed for decades, but never before has there been a moment when mass market applications were truly within reach. Devices in the 1990's were prohibitively expensive for all but the most well-funded research labs. Even for those who could afford them, the infrastructure and ecosystem to maximize the technology's potential was lacking. The internet was nascent and not yet able to support multi-user interaction, and there was not yet a global community of 3D artists and developers to provide a wealth of high fidelity digital assets and data. Today, the situation is different. Multiple six degree-of-freedom virtual reality systems have been released at consumer prices in the past two years, creating a broad user base. Developers are familiar with 3D authoring workflows from their experience in the lucrative PC gaming industry. The internet now allows people to connect with one another in VR, and plays host to an ecosystem of resources and tools for content creation.

The face-value reason to consider virtual reality for learning is that it allows people to have experiences that are relevant and highly instructive, but difficult or impossible to access in the physical world. Learners can experience what it's like to be on the moon or stand inside a living cell. They can operate expensive equipment without worrying about making mistakes. They can interact with people from faraway places, and perform super-human feats. Annotations and contextual information can be embedded directly into the world and presented in-situ at relevant moments. These are all novel prospects, but in a world where learning practices are entrenched and slow to change, we must ask: do the advantages justify the effort required to change the system? To break it down one level further, I ask the following questions:

- What are the advantages of VR as a medium for learning?

- What hurdles exist to incorporating VR into practice in real-world settings?

- What affordances and interaction techniques are required?

- What advantages can be shown empirically?

This dissertation represents an effort to investigate and construct answers to these questions. The

background (Chapter 2) provides a theoretical basis for claims about the advantages of VR. Chapter 3 presents the design framework of the *Equipped Explorer*, which categorizes the different purposes that tools and environmental affordances fulfill from the perspective of the learner. This lays the groundwork to survey these comprehensively through the building and testing of prototypes. Then it states some of today's challenges to integrating VR into real-world contexts, leading to the concept of *Reality Integration*, which addresses these problems. Chapter 4 presents numerous prototypes and design explorations that attempt to provide insight into all of the above questions, framing the discussion in terms of the Equipped Explorer framework and the concept of Reality Integration. Then, Chapter 5 presents two experiments I carried out, taking an empirical approach to the question of how VR compares with existing media, considering first learning abstract concepts and then communication and collaboration. Finally, in Chapter 6, I conclude with thoughts on the implications of this dissertation. One key idea to explore moving forward is the use of VR to encourage reflection in the learning process. Another is empowering learners and educators with tools for creating interactive learning experiences in VR.

Overview: Background

The background section is broken into four parts: I position my treatment of exploratory learning, present relevant concepts from cognitive science, discuss research on the use of virtual labs and simulations for learning, and then review related work on VR for learning and training. Exploratory learning belongs to the tradition of constructivism, and I summarize the history of constructivism, including some instructional approaches it has given rise to. I introduce concepts from cognitive science that provide reasons to suppose that VR could yield advantages in the area of learning. These are *embodiment*, *spatial learning*, *cognitive load*, *learning with multiple representations*, and *note-taking*. Embodiment is relevant on the one hand in the context of *embodied cognition*, which describes the connections in the brain between representations of body movement and those of abstract thought [Wil02a]. On the other hand, embodiment refers to the property of a human-computer interface in which the body is represented directly in space [Dou04]. Next, spatial learning and spatial intelligence refer to a set of measurable competencies involving mental manipulations in 3D, which correlate with academic performance in math, science, and social studies. Cognitive load theory seeks to characterize the most salient aspects of information presentation that make given

21

material easier or harder to assimilate. The study of learning with multiple representations has un-covered both the necessity and the pitfalls of using multiple spatial and semantic representations of systems or concepts. Finally, note-taking plays an important role in learning. It can be under-stood in terms of contextual memory and cognitive load, and there is related work that illuminates the relevant mechanisms and considerations. Next, moving on to related work, research on vir-tual reality for learning and training has focused on medical applications, military and aerospace, scientific visualization, and science education.

Overview: Design Framework for Learning in Virtual Reality

This section provides a method for describing and prescribing the affordances of VR learning appli-cations, considering first the in-VR experience using the Equipped Explorer framework, and then its relationship to the world outside VR, introducing the concept of Reality Integration. The Equipped Explorer design framework focuses on two components or layers of consideration: the *activities and tools*, and the *medium*. The four categories of activities and tools correspond roughly to *ex-ploration, collaboration, capture*, and *review*. Next, I introduce the discussion of the medium by presenting challenges to integrating VR learning experiences in real-world contexts today. This motivates the definition of a concept of *Reality Integration*, which delineates a set of medium affor-dances that will solve the most important problems. Broadly speaking, it concerns the accessibility of content across all digital platforms inside and outside what we now think about as VR. The goal is to support individual use across contexts, along with the critical and fundamental *social* usage of the medium.

Overview: Prototypes and Tools to Equip the Explorer

In this section I present *prototypes*, which are applications built around particular content, appli-cations, and models of interaction, and *tools*, which are generic, reusable mechanisms, involving hardware, software, or both, for VR interaction in the context of different applications and pro-totypes. These explorations are broken down into three categories: single-user, multi-user, and Reality Integration. The single-user prototypes are used to pilot and uncover tools that facilitate the Equipped Explorer's activities of exploration, capture, and review. The first one explores how free-hand input methods can be used to explore physical simulations. The examples are a retinal circuit,

a viral infection, and electrostatic dynamics. The next set of prototypes explores bookmarking and annotation as a form of active capture or note taking. The next category of prototypes concerns multi-user applications, discovering needs and piloting tools for communication and collaboration. I build around a collocated user scenario, with multiple users' physical and virtual worlds aligned. Through an initial exploration of sketching together in 3D, I discover user needs associated with in-world creativity and creation, along with environment exploration. This leads to a set of questions around *Reality Integration*, that is, making practical use of work done in VR and connecting the worlds inside and outside the headset. Besides the tools, another outcome of this chapter is a pair of focused questions that are addressed through experimental studies in the next chapter.

Overview: Experimental Studies

Two specific questions arose in my work on prototypes and tools for learning in VR: (1) is there a fundamental learning advantage to embodied VR interaction, in terms of comprehension, recall, and retention? (2) can minimal embodied avatars in VR provide an adequate or even advantageous basis for learning-focused collaboration? In this chapter I present two experiments: one related to the first question, and one related to the second. The first experiment concerns comprehension and retention in VR using electrostatics as a topic area, and compares learning using 2D vs VR interfaces. The results show promise for learning in VR, but also demonstrate that advantages are obtained only when applications really utilize embodied interaction and the third dimension. The second experiment explores communication in VR, comparing face-to-face and VR interactions side-by-side. Participants play word guessing games, and the experiences are contrasted using questionnaires, interviews, and physiological feedback. The results show that VR is similarly effective, so long as the avatar is not missing a particular reference point on the body that is critical to the use case. I also uncover certain advantages to VR communication over face-to-face communication.

Overview: Conclusion

The work presented in this dissertation, including a theoretical framework for exploratory interaction and Reality Integration, a set of learning-focused tools for VR interaction, and a set of experiments proving out the usefulness of VR for learning, suggests that there are valid reasons to pursue virtual

reality as a medium for learning. These reasons include measurable learning gains while using VR versus 2D interfaces, certain advantages of communicating in VR versus face-to-face, and evidence that VR can be integrated into existing contexts in a seamless way. With the learnings from the prototypes and experiments in mind, I propose leveraging VR's ability to record and re-play as a foundation for two different but equally important purposes: encouraging and supporting the process of reflection during learning, and empowering learners and educators with tools for authorship. I then summarize the most important questions that I see moving forward: (1) When is VR advantageous, and how can we focus content development around these areas? (2) How can VR be integrated into learning practices, at physical institutions and online? (3) How can we enable instructors and students to create content without requiring a software developer?

Chapter 2

Background and Related Work

Before introducing my design framework in Chapter 3, I will present background and related work to highlight opportunities and possible advantages for learning in VR. In Section 2.1, I frame exploratory learning in the constructivist tradition, surveying related concepts in the process. Then, in Section 2.2, I review concepts from cognitive science that may be leveraged to yield learning advantages: embodiment, spatial learning, cognitive load, learning from multiple representations, and note-taking. In Section 2.3 I look at how exploratory learning has been implemented so far in the context of simulations and virtual labs. Finally, Section 2.4 presents related work in VR. This is divided into subsections on *training*, for example in medical and military applications, and *learning and collaboration*.

2.1 Constructivism and Exploratory Learning

In this section, I discuss *constructivism*, and position exploratory learning as one of several related learning paradigms.

Constructivism is a philosophical paradigm often applied to learning theory, which holds that people do not "absorb" knowledge, but rather make meaning through interaction. John Dewey is generally credited with introducing the notion in 1933 [Dew33] and it is contrasted with the conventional learning paradigm based on passively listening to lessons and lectures. Jean Piaget advanced this theory, exploring and providing evidence from developmental psychology. His numerous ex-

periments provide crisp insights into the development of cognition, and the way that knowledge is acquired through interaction with the world [Pia52]. Jerome Bruner's work took place in parallel, arriving at similar conclusions about the active nature of learning, but disagreeing on the causal relationship between the development of language versus cognition. Bruner's work is also focused more on how learning occurs in context, including the involvement of experts, and considerations about how learners go "beyond the information given" when they acquire and master concepts [BA73].

Numerous learning theories have built on constructivism, emphasizing aspects of the meaning-making process that are applicable in particular settings or for particular kinds of subject matter. The notion of *situated learning* focuses on the passing on and learning of craft. It was established by Lave and Wenger [LW91], and centers around the ideas of *communities of practice* and *legitimate peripheral participation*. A related body of research is developed by Rogoff, et al. around *guided participation, intent community participation*, and ultimately *learning by observing and pitching in*, which concerns itself with the passing on of cultural traditions in indigenous communities [RMG$^+$93, RPA$^+$03, PR09].

Kolb's theory of *experiential learning* describes a cycle involving experience, reflection, abstraction, and active experimentation. This applies to learning in environments that afford a rich set of interactions, and where the learner has autonomy [Kol84].

Another notable offshoot of constructivism is *constructionism* [PH91]. Constructionism uses the creation of artefacts as a central component of constructivist learning. Artefacts may be computer programs or the output they produce, such as with the use of Logo, Starlogo, or Scratch [Pap72, CKR01, MBK$^+$04], or physical artefacts produced by construction kits like Lego Mindstorms (one early study of the use of MindStorms in an educational setting is Kassner [Kla02]). The definition of the desired outcome by the learner, and the subsequent self-validation of achieving that goal are core components of this learning theory.

The characteristic of learner autonomy is one of the most central to *exploratory learning*. In its original sense, exploratory learning referred to an alternative method for learning to use computer interfaces. The term was coined by Rieman [Rie96], but the author points to earlier work

[Mal82, Car82, Shn83] that had begun to investigate explicitly designing computer systems to encourage exploration, using fun and pleasure to assist in onerous learning tasks. The term gradually broadened in scope to eventually refer to learning any topic in a learning environment that gives the learner the autonomy to choose what to interact with and how to interact with it, although it does always seem to refer to settings that are computer-mediated [MGSGN13].

This style of learning assumes that the environment affords a rich set of interactions, as with experiential learning. The notion of the "experience" phase of learning in Kolb's cycle seems to assume the existence of things happening that can be experienced passively, while exploratory learning appears to grow out of a notion more akin to a sandbox, where nothing is happening but that which the user deliberately set in motion. These are not mutually exclusive, nor are they strict, but I claim that they fairly characterize the origins of these respective conceptions of constructivist learning.

From here on, I will talk about *exploratory learning*, since it is best aligned with the unique advantages of VR that I will present. For the moment suffice it to say, very broadly, that VR differentiates itself from mere viewing of 3D displays most powerfully in the direct control afforded by embodied spatial input through the tracking of the head and the hands in space. Therefore, it is through forms of learning that put the user in control where we can expect to reap the advantages of VR. Note, however, that other theories like experiential learning and constructionism translate just as readily into VR.

2.2   Cognitive Aspects of Information Presentation and Interaction

In this section I present related work that concerns the presentation of information, and how we interact with it. First I will introduce the notion of embodiment, including embodied cognition and embodiment in interaction. Next I will talk about spatial learning and how it is important especially for science and mathematics. I will introduce the theory of cognitive load and the model of working memory. Finally, I will discuss the theory of learning from multiple representations. In each case, I will highlight possible advantages of learning in VR that can be deduced from the respective theory.

27

2.2.1 Embodiment

There are two aspects of *embodiment* that are relevant to VR. The first is a style of human-computer interaction, and the second is a theory about cognition (*embodied cognition*).

In human-computer interaction, embodiment refers to the property of an interface where the motions of the body in space are directly mapped to the motions of a component in a spatial interactive system [Dou04]. For example, given a task like placing virtual objects on a simulated shelf, a non-embodied interface might use a joystick, and an embodied interface would be a VR glove which allows the user to grab the relevant objects and manipulate them to the shelf directly in space. This kind of interface may derive efficiency advantages [BKJJLP01], although not without addressing challenges related to learnability [HPGK94]. It has also been extensively explored in connection with the notion of *presence* [SUS94, Sla99, KGS12, CB16]. In the field of psychology, presence has been shown to influence user's choices in role-playing scenarios [BB, BLB+02, Bla02, LBB99].

Embodied cognition is a subject of research exploring connections in the brain between abstract concepts and motor representations of the body interacting with the world [Wil02b]. Imaging technologies have demonstrated that these connections exist, showing activation in the same brain regions when planning or carrying out physical activities, and when performing abstract thinking tasks [MC08]. The debate rages on about how to interpret these connections, in particular to what degree the core of abstract concepts are inextricably tied to motor regions, or are just activated when dealing with those concepts which are represented independently. Either way, learning abstract concepts in connection with body motion, as is done in an embodied VR interaction, should provide an advantage over alternatives that do not use body motion.

2.2.2 Spatial Learning

Spatial intelligence and spatial learning have been investigated as easily measurable correlates with achievement in math, science, and social studies [New13]. A typical spatial intelligence task involves performing mental rotations of 3D objects represented in 2D on paper or on screen (example in Figure 2-1). Evidence suggests that these skills can be trained and the benefits of do-
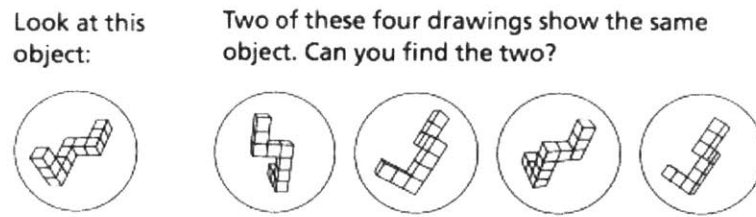
Figure 2-1: Mental Rotation Task [VK78]

ing so transfer to the areas mentioned above [New13]. It seems plausible that practicing spatial intelligence tasks in VR (that is, in 3D and using 3D interaction) could accelerate this kind of training.

2.2.3 Cognitive Load Theory

Cognitive load theory concerns itself with understanding what aspects of information presentation make subject matter easier or harder to learn. The underlying assumption about cognition is the concept of *working memory*– the capacity for mental effort at a given time– as a limited resource. The *cognitive load* associated with a task is the amount of this resource required. A common model of working memory involves several components [Bad07], and the ability for multiple tasks to be performed or managed simultaneously depends not just on the overall cognitive load of each, but also which components of working memory each requires. Therefore, in order to design for learning processes that involve multiple concurrent tasks, understanding the utilization of the components of working memory is critical.

The four components of the model of working memory mentioned above are the *central executive*, the *phonological loop*, the *visuospatial sketchpad*, and the *episodic buffer*. The central executive is the controller of attention; it doesn't have any storage of its own, it just controls the allocation of attention. The phonological loop is a memory for sound. It fades quickly unless it is rehearsed, which can be done with the help of the central executive. The visuospatial sketchpad is a perceptual memory. A rehearsal mechanism for the visuospatial sketchpad has been proposed, which involves eye movement [AJRI98]. The episodic buffer, is an interface between the aforementioned components of working memory and the long-term memory. The episodic buffer is important for explaining how long-term memory is recalled into working memory and how the phonological loop

29
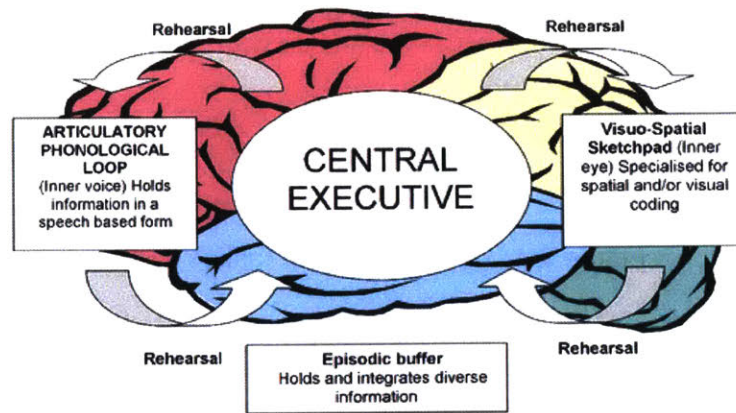
Figure 2-2: Baddeley's Work Memory Model

and visuospatial sketchpad give rise to higher-level episodic understanding which can then be stored in long-term memory. This four-part model is pictured in Figure 2-2.

Building on this theory of working memory, cognitive load theory identifies properties and effects related to learning materials that make them easier or harder to digest. Different teaching materials are tested and evaluated on learners' retention and comprehension [SAK11]. The most important overarching concept is that of *intrinsic* versus *extraneous* cognitive load. Intrinsic cognitive load refers to the unavoidable and inherent challenging aspects of learning content. Extraneous cognitive load refers to challenges in comprehending learning materials that result from how they are represented rather than the content itself. The advantage of VR, from a cognitive load perspective, is that it opens up new avenues for decreasing extraneous cognitive load.

There are two subtly different aspects of using 3D space in VR that I claim can decrease extraneous cognitive load. The first is related to the *split-attention effect*. This effect occurs when written or visual information is presented in such a way that it is split spatially between parts that need to be integrated in order to be understood. The requirement to look quickly back and forth between disparate regions incurs extraneous cognitive load that impedes learning. An example with and without the detrimental design aspect is shown in Figure 2-3. Using 3D space and responding to the current focus of attention, VR allows semantic information to be situated at its spatial reference point, which should allow the split attention effect to be mostly eliminated. The second aspect relates to translating between 2D and 3D representations. Even with careful attention to

30

graphic design, representing 3D phenomena in 2D requires the learner to perform a mental translation, rendering phenomena in 3D in the visuospatial sketchpad based on a representation in a 2D medium. Representing 3D phenomena in 3D, combined with the natural perspective selection interface afforded by hand and head motion in VR should decrease the extraneous cognitive load associated with this rendering process.

a Find ∠ EFB
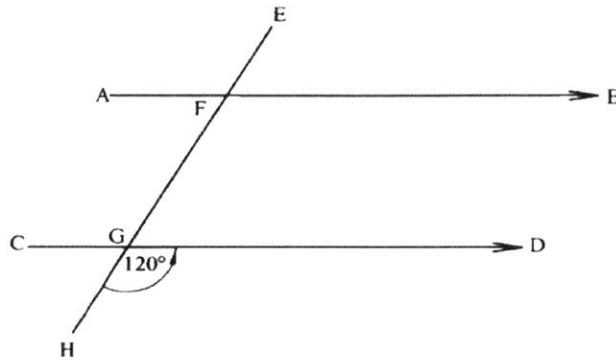
∠ GFB = 120° (corresponding angles between parallel lines AB & CD)

∠ EFB = 60° (angles on a straight line sum to 180°)

b Find ∠ EFB

Step 2: ∠EFB = 60° (Angles on a straight line sum to 180°)

Step 1: ∠GFB = 120° (Corresponding angles between parallel lines AB
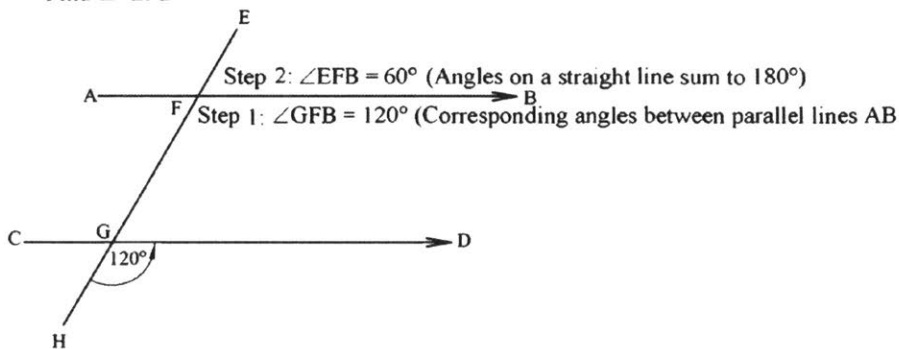
Figure 2-3: Split Attention Effect [SAK11]

2.2.4 Learning with Multiple Representations

Physical systems are almost always considered from multiple perspectives, whether it be their physical structure, behavior over time, contrasting characteristics before or after a process occurs, or others. How learners connect such representations has been a topic of research, existing at the intersection between cognitive science and the learning sciences. The split attention effect is
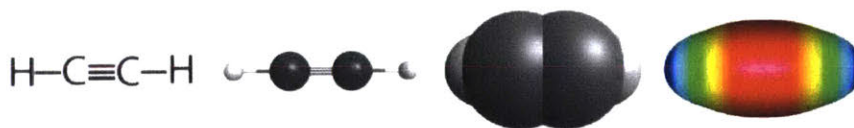
Figure 2-4: Example of Multiple Graphical Representations in Chemistry [RMF15a]

related to the issues confronted in this area, but the theory of learning from multiple representations (MRs) delves somewhat deeper into the usage of materials as part of an instructional strategy than does cognitive load theory.

Two branches of literature I'll mention refer to learning from *multiple external representations* (MERs) [Ain06], and *multiple graphical representations* (MGRs) [RAR15, RMF15a]. The former has led to a set of design heuristics for learning materials and instructional approaches. For example, two such heuristics are: (1) learners must understand each representation in isolation before they can successfully connect multiple representations, (2) constrained representations that represent a subset of more complex representations can be very successful, assuming that the relationship between the two can be made clear (e.g. by presenting them at the same time).

Literature on learning with MGRs has focused more explicitly on integrating strategies that structure the learning process. One example is introducing self-explanation prompts [RAR15], and another is using intelligent tutoring systems (ITS's) [RMF15a], that can consistently and automatically control how and when other representations are introduced. An example of multiple graphical representations is shown in Figure 2-4.

The promise of VR from the perspective of learning from MRs is its ability to connect and merge such representations. Overlaying multiple spatial representations, for example, will allow 3D space to act as a consistent point of reference while switching between the representations. As a simple example, connecting orthogonal projections (e.g. onto the XY and XZ planes) can be done in place, showing them situated in 3D space. In VR, the user can use head motion to alternate between the perspectives, and thus connect them in a natural way. Secondly, of course, VR affords the same abilities as 2D interfaces in terms of implementing ITS's, but with additional inputs to respond to. For example, a VR ITS can respond to head movement and the taking of particular perspectives, in order to actively highlight visual phenomena that become uniquely clear in each case.

2.2.5 Note-Taking

Note-taking is a powerful tool for enhancing the effectiveness of learning activities, which goes back at least to ancient Greece, where the early form of the notebook was known as a *hypomnema*. Two functions performed by note-taking are (1) capturing the information that a learner is exposed to (*external storage*), which allows the information to be reviewed later, and (2) facilitating deep understanding through paraphrase, summarization, and so on (*encoding*) [POK05]. Due to the limitations in *working memory* [Bad92], it has been observed that note-taking imposes a tradeoff between production and comprehension – the more time and attention that is devoted to the writing of notes, the less there is to devote to understanding the content [POK05]. When more time is devoted to production, notes tend to be "verbatim," and this style of notes is known as *non-generative* since it does not require or reflect that the learner has understood the material, whereas when more time is devoted to comprehension, notes can be *generative*, capturing the output of a process of idea synthesis [MO14].

Formal studies of the impact of note-taking disagree on its learning value. As highlighted by Lin and Bigenho [LB15], variations may be explained by differences in cognitive load associated with the particular systems and content in question. This is supported by their study showing that introducing distractions changes which of several note-taking methods yields the best learning outcomes [LB11]. A recent study that pitted an HMD-based VR learning system against a slide-show based learning system on a 2D display found the latter to be more effective, noting that note-taking was only possible in the non-VR system [Lom14]. Taken together, a valid hypothesis remains that, with careful attention to the cognitive load imposed both by the environment and system affordances, it is possible to design learning systems (e.g. in VR) with support for note-taking that yield better learning outcomes than their counterparts without note-taking. One concrete exploration of this idea will be explored in Section 4.1.3.

2.3 Exploratory Learning with Simulations and Virtual Labs

I mentioned in Section 2.1 that the notion of exploratory learning originates from the idea of learning a computer interface through exploration, as an alternative to the once ubiquitous paper manuals that are virtually nonexistent today. Very early on, however, researchers began to investigate the

idea of learning about the world (e.g. scientific or engineering principles) through exploration using a computer interface. In this section, I will discuss the development of exploratory learning from computer simulations and virtual labs.

One of the first examples of research on learning from simulations in higher education used the domain of control theory [NDJ93]. Students were encouraged through the use of a paper worksheet to go through a scientific process of hypothesis forming and testing. The results indicated that students were more successful at learning from this process when they were given hypotheses to test. A recent review including this and much subsequent work on virtual labs indicated the proper use of guidance in the process of exploration as the single most important factor in their successful application [dJLZ13]. This same review also concluded that virtual and physical labs can be complementary, with virtual labs sometimes being preferred since they allow more time and attention to be focused on the sense-making process of data analysis than fiddling with the data-collection apparatus (the latter is sometimes valuable, but sometimes perceived as tedious and irrelevant).

The possible advantages of VR when it comes to exploratory learning from simulation and virtual labs lies in the ability to connect and merge representations and eliminate extraneous cognitive load. Virtual labs are used to connect representations as well, but VR can do this even more effectively. The example in Figure 2-5 shows a lab bench equipped with a 3D configuration of physical objects and light sources and a projection screen at the end, along with a 2D representation of the light rays on the wall adjacent to the bench. Assuming that the virtual lab can be inspected from different locations and perspectives using the desktop computer interface, the alignment of the 2D representation with the bench to allows the learner to more easily interpret the information it gives about the system, when compared with a textbook representation.

Even so, the virtual lab creates a split attention effect by using two spatially separate graphical representations of the same system. This choice was most likely made since on a desktop display, overlaying the 2D rays on the experimental apparatus, or even drawing 3D rays in place, would create a cluttered visual representation. In VR, the two representations could be overlaid in place, eliminating the split attention effect while keep the information easily interpretable. This is because
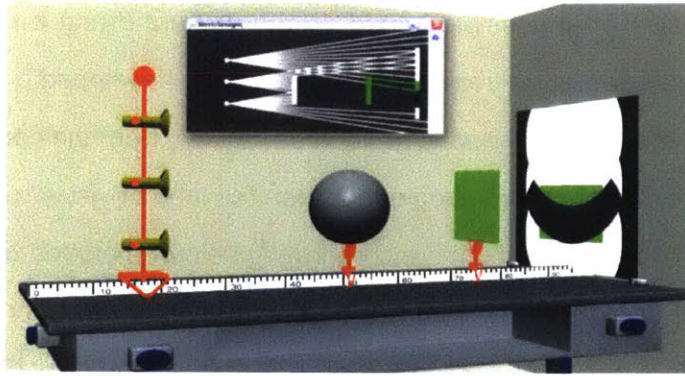
Figure 2-5: Virtual lab that connects multiple representations [dJLZ13]

the use of head movement afforded by VR allows learners to quickly and intuitively survey the system from different perspectives, distinguishing and relating the spatially coincident information–in this case, the rays that fill the 3D space, and the objects that they intersect.

2.4   Related Work

In this section, I will review related VR work in three categories: medical and military training, scientific visualization, and learning or education. I will highlight learnings that are relevant to the work in this dissertation.

2.4.1   VR for Military and Medical Applications

The applications of VR (and augmented reality) I'll discuss in this section have one thing in common: it made economic sense to explore them at a time when the necessary equipment was prohibitively expensive for most applications. I will mention the particular reasons why as I introduce them.

One of these high-stakes use cases is medical training. Being able to practice surgery in an environment free of consequences is something that isn't possible any other way. Furthermore, with human life on the line, even the great cost of the equipment would be worth it in the case that advantages were borne out in practice. The growth of VR in the surgical training space was reviewed in 1999 [PAT99], concluding that the application was sure to grow.

Another family of applications from the previous wave of VR was military and aerospace. One advantage in this case is the operation of expensive equipment in a consequence-free environment. In some cases it also helps deal with the scarcity of the said equipment: spending many hours

to train is either expensive or impossible based on the amount of equipment available and the number of operators that need to be trained. The US Airforce developed a set of virtual and augmented reality systems that worked hand in hand, most notably from the early 1970's until the late 1980's, proving out a set of ground-breaking principles. The idea of *visually coupled systems* goes back to at least 1974, referring to systems incorporating helmet-mounted displays and positional tracking of the head. These quickly gave rise to the need to integrate such systems into training simulators, leading to the Visually Coupled Airborne Systems Simulator [BF74, Haa84, Koc77], and ultimately the Super Cockpit. The Super Cockpit was an augmented reality system designed for and by the US Airforce to improve the performance of fighter pilots [TAF86]. The argument was that the presentation of information was cognitively taxing, and the input methods unnatural. Augmented reality could better leverage the numerous streams of sensor data available to such a pilot by condensing them and presenting only the cognitively salient aspects of the information. Simultaneously, innovations were made in the virtual reality counterparts of this system, proposing ways of using virtual space to design training scenarios [FI87].

Taken as a whole, this work shows that, even with the comparatively cumbersome hardware of the day, simulating reality can have real benefits in terms of both how information is presented, and how learners can be given access to otherwise hardly accessible experiences.

2.4.2 VR for Scientific Visualization

Numerous publications appeared in the 1990's concerning the use of VR for scientific visualization. One discussed the use of VR with haptic displays, using as a test scenario chemists identifying docking positions of molecules [BOYBJK90]. Another reviewed the benefits and technical challenges. It cites intuitive interactions with complex datasets as an advantages, and also highlights that representing data is low-hanging fruit compared with photorealistic renderings of environments, which were at that point out of reach [Bry96]. Some examples of other projects are a VR wind tunnel , medical data visualization, and a visualization related to general relativity [BL91, BFO92, Bry92]. This related work shows that there is great promise in continuing to experiment with the use of immersive, interactive interfaces with the goal of gaining deeper understanding of complex data and systems.

2.4.3 Learning and Education

As long ago as 1972, some people began to take seriously the idea that computers could transform education [Kay72], calling it "science fiction" while stating that the trajectory of technology would "almost guarantee" it to happen. By 1995 the vision of immersive, simulation-based learning environments that would allow learners to interact with local and remote others in real-time was clearly articulated [Ded95]. The ensuing research focused in particular on the ability of these kinds of environments to facilitate constructivist learning [DSL96, ROS06]. A review of this work from 2009 concludes that the most salient strengths of immersive virtual environments for learning are their ability to (1) show multiple perspectives, (2) facilitate situated learning, and (3) transfer to the real world [Ded09]. Given the sparsity of prior work, it seems clear that further exploration is highly justified, and should happen in a variety of subject areas in order to better characterize the unique advantages offered by VR. In this dissertation, I survey categories of interaction that are relevant to learning by developing numerous prototypes, and delve into the question of the unique advantages of VR with two experimental studies.

2.4.4 Discussion

For each area of related work in this section, I drew conclusions relevant to the application of VR to learning. I would now like to put these conclusions into historical perspective. Firstly, the application to elementary and higher education was difficult to justify from a resource perspective at the point when the equipment was very expensive. Even so, some of the publications cited above (most notably ScienceSpace) do represent pioneering work. Now the situation is different with drastically lower prices per device and the existence of the internet making the vision of ubiquitous devices, remote interaction, and distributed content creation much more in reach. Second, the related work has demonstrated that the general approach of using VR for learning is promising, but today in some cases we might obtain different results using the current generation of less cumbersome and higher performance hardware. It seems clear that a great many questions remain to be explored in terms of what forms of interaction and visualization are most beneficial to learning, for what subject matter these are most applicable, and how to begin designing for VR as a comprehensive learning platform as opposed to an occasional supplement to traditional learning practices.

Chapter 3

Design Framework for Learning in Virtual Reality

In this chapter, I develop the theoretical underpinnings of the rest of the thesis; providing a backdrop for the design explorations and experimental studies. The broad notion is that the learner should be thought of as an explorer or a pioneer in the wilderness. She is equipped with tools for a variety of tasks, from survival to experimentation and documentation. As she explores, she stores objects she discovers and wishes to keep for future use or reference. In the first section, I lay out the critical categories of purposes and interactions that an explorer can have as part of her experience, and use this as a framework for understanding what sort of user interface is needed to support a wide variety of learning and discovery processes. In the section that follows, I survey the landscape of existing technologies through the lens of these capabilities, to establish a context for the great opportunities that virtual reality (or more generally, *virtuality*) offers. Finally, I examine the foundation of this set of experiences, and propose the notion of *Reality Integration*, casting our present conception of "virtual reality" as only one component in a more comprehensive infrastructure required to make virtuality into a medium that is maximally usable and impactful.
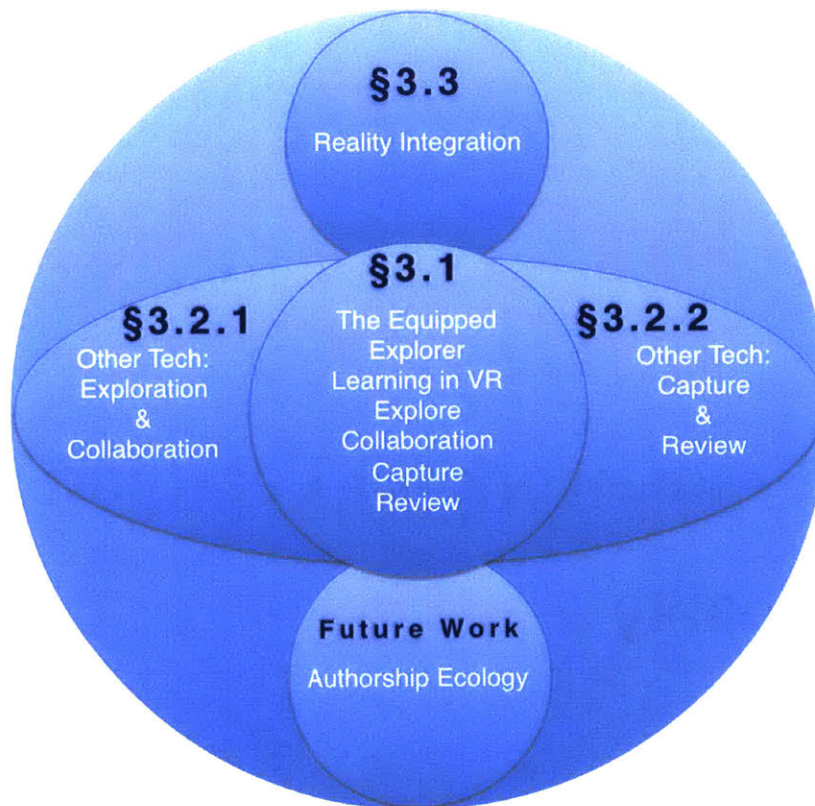
Figure 3-1: Overview of Design Framework Chapter

3.1 The Equipped Explorer: Tools for Learning

As mentioned in the background section, I adopt the paradigm of exploratory learning and the capability or aspiration for self-regulated learning as a foundation for this theory. The learner is presumed to be both **active** in drawing conclusions from her experience, and **purposeful** in her use of tools to capture experiences for later review. Under these assumptions, my goal here is to approach the support of learning from the standpoint of technological affordances. That is, how does one categorize and frame the wide variety of tools and experiences that may be made available to learners? I present the notion of the *Equipped Explorer* as an empowered, autonomous participant in an environment to be explored, experienced, and experimented with. This includes an *a priori* tool taxonomy which categorizes tools in terms of their purpose, modality, and scope. This taxonomy may be used either descriptively when studying systems or prescriptively when designing them.

In this framework, a learner is immersed in an interactive environment, and the assumption is that

she has a conception of purpose. The purpose may be very broadly or vaguely defined, but interaction is driven forward by her general notion of aiming to explore, understand, or master something associated with the environment. The framework has the dual purpose to describe the tools in relation to the environment, and the environment in relation to the tools – since in general both of these are encompassed in the design of an experience. What separates the two, concretely, is the perceived agency behind changes or occurrences in the experience – if the explorer "causes" something, then she is employing a tool; whereas if the ostensible "cause" comes from the environment, then it is part of the world and is outside of her control. The experience developer, of course, has equal ability to shape both tools and the environment, and must consider how the two relate to each other. That being said, there is a long tradition of developing immersive and non-immersive environments, and what is unique about the Equipped Explorer framework is primarily its consideration of contextual tools for learning and discovery. Hence, although there are many ways to expound on design considerations for the interactive world, it is the explorer's "equipment" and the attributes of the environment that define or enable its functioning that will be the topic in this chapter. In this section, I elaborate the elements of this vision: I first present a taxonomy for tools, in order to be able to classify and compare them. Then I discuss tool ergonomics, to provide a language for analyzing choices about how tools are accessed. Finally, I present a basic reference set of tools that fulfill the most generic learner needs within the present set of assumptions, specifying their classification and recommended ergonomics.

3.1.1 Taxonomy of Tools

My taxonomy of tools provides three dimensions in which tools are assigned attributes: (i) the **purpose** of the tool; (ii) the **input modality** required to employ the tool; and (iii) the **scope** of the tool, which relates to how frequently the tool is used, and in what different contexts.

Purpose

Recall that the metaphor for the theory and tool taxonomy is as follows: the Equipped Explorer travels through the world, and carries on her person numerous tools, as well as a sort of storage receptacle, for things that she wishes keep. She deliberately exposes herself to content through exploration and experimentation, mentally and physically taking note of insights and observations,

41

and reviewing these whenever opportune. In virtual reality, the tools she carries are related to (i) communication with other people, (ii) environmental and object-based interactions, including locomotion and manipulation, (iii) the act of selecting and preparing things to be carried in the backpack, and (iv) arranging the backpack to allow things to be found and retrieved quickly. Note that any or all categories of purpose may occur simultaneously. For example, a learner seeing something new, glancing at notes from earlier activities, and sharing her resultant observations with another explorer. These four *purposes* in the tool taxonomy, referred to as *communication, exposure, capture*, and *review*, are depicted in Figure 3-2. The categories are further subdivided, and some basic tools for each are enumerated, though the lists are not meant to be exhaustive.
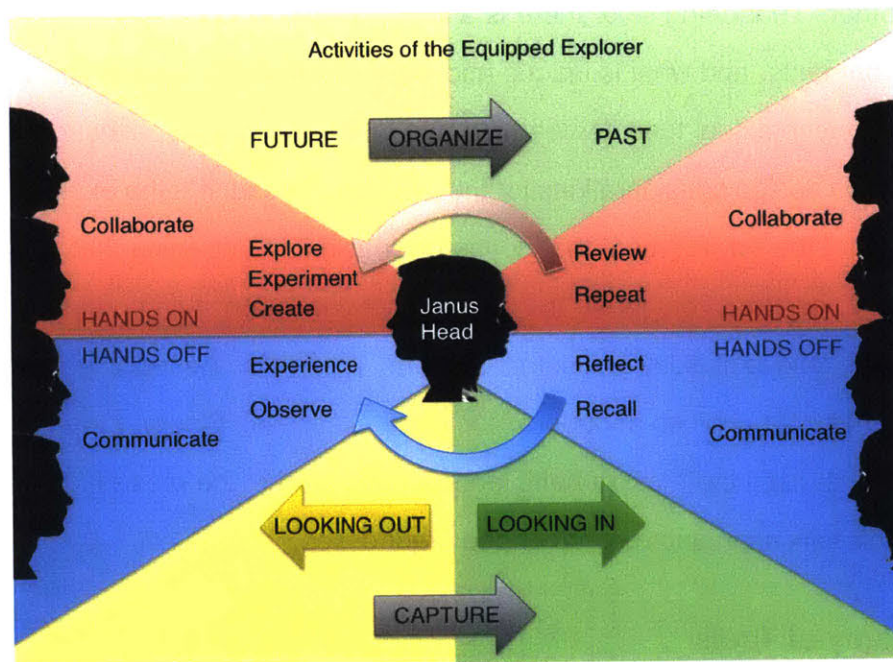
Figure 3-2: Schematic of Interactions Between Explorer and Virtuality

Explore   These tools are aimed at seeking out, engaging with, and regulating activities in a new or external environment. I divide these into the categories of *explore, experience (v.)*, and *experiment*. Other categories may be possible, but these are sufficient to provide tool categories in the scope of this theory.

Unlike the wilderness that real-world pioneers or explorers have explored, the environment our Equipped Explorers are immersed in is also part of our design. That is, a VR learning experience designer has the two separate but intertwined tasks: (i) to design the tools that the explorer per-
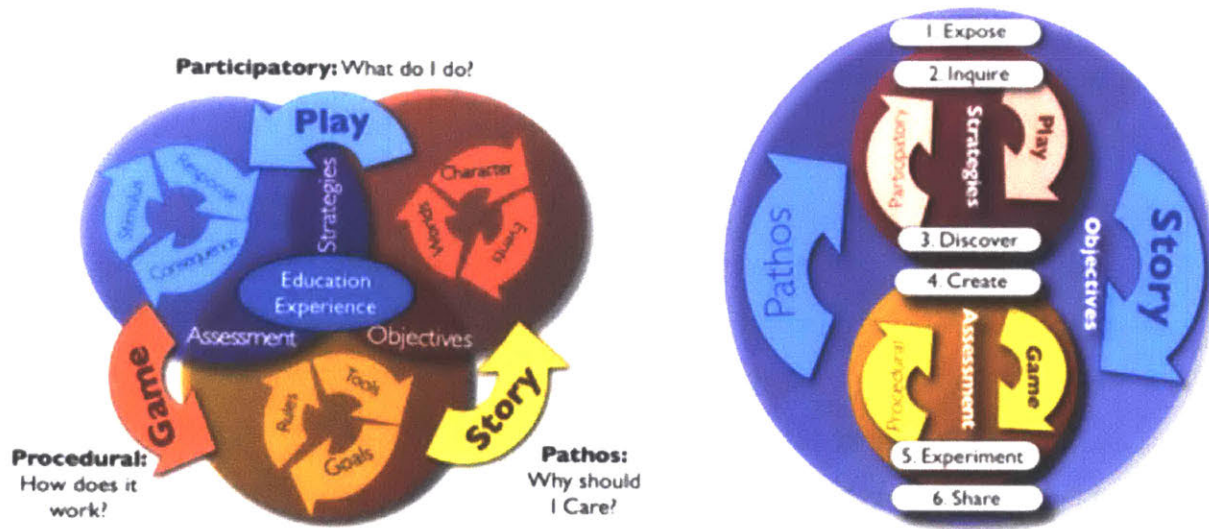
Figure 3-3: Interplay Instructional Design: Motifs (left) and Strategy (right) [SCH12]

ceives as an extension of his/her self, which must be consistent and quick to access and operate (I refer to as the tool belt), and (ii) to design the environment, which, from a technical or implementation perspective, is just as much a part of the software as the tool belt, but is conceived of and perceived by the user in an entirely different way, as being fundamentally "outside the self."

The InterPLAY instructional design paradigm [SCH12] applies principles of interactive entertainment to the use case of learning, and proposes the interaction motifs of *story*, *play*, and *game*. These establish a cycle of engagement driven by the questions "why do I care?", "what do I do?", and "how does it work?". The instructional strategy builds on these motifs to create an explicit set of steps that can be supported through design of learning materials. The conventions and strategy are shown in Figure 3-3. Inspired by InterPLAY, I propose three learner-centric environmental affordance and interaction categories: *experience*, *explore*, and *experiment*. Abstractions and tools that implement these categories can be used as a basis for exploratory learning or related approaches.

- **Explore**. Basic exploration affordances include: teleport/locomotion, scaling, browsing affordances, information scent (thumbnail/preview), and spatial map abstractions.

- **Experience**. I use *experience* as distinct from *explore* to refer to the passive nature of the user's relationship to things that are "happening" in the environment. This does not preclude

43

interactivity, but sets apart "Learning by Observing and Pitching In" [RMG+93] or "Legitimate Peripheral Participation" [LW91] from sandbox-like settings where the user is directly responsible for driving progress forward through interaction.

- **Experiment**. I refer to experimentation behavior as carrying out a multi-step constructive procedure with the goal of answering a question. For example, moving a charge around continuously and observing how the electric field varies would qualify as exploration but not experimentation (despite the fact that this might fall under some colloquial uses of the latter). Varying the altitude angle of a canon and measuring corresponding variations in distance projectiles travel would be an experiment. Affordances that make experimentation possible fall in both "body-based affordance" and Experiment affordance categories. For example, it may be that the Equipped Explorer always has a measuring tape, because this can be used in almost any situation to quantify aspects of the environment, while the Geiger counter that might be appropriate in a virtual radioactive environment would be more likely to fall into the instrumentation specific to the place or scenario. (body-centric versus context-centric experimentation affordances).

Collaborate   These tools are aimed at communication, coordination, and other affordances relevant to collaboration.

A minimal avatar, representing the headset and handheld controllers, is sufficient to convey co-presence – including awareness of visual attention, and partial information about comprehension and affect. As such, the avatar itself is a communication tool employed by one explorer to communicate with another explorer. Less minimal avatars – representing for example facial expressions and other parts of the body, may increase the expressive power. If they are not operated "directly" as their physical counterparts would be (e.g. using buttons to change facial expressions), care should be taken, since this may distract the avatar user, and create unnatural-looking signals for their communication partner.

Hands that are part of a minimal avatar can be used to gesture expressively, and also to point, to indicate a focus of attention in the shared environment. It should be noted that using hands to point, while very natural to use, does have natural shortcomings that can be improved upon in a virtual

environment. In fact, if we abstract and extrapolate the "pointing" functionality, we can arrive at an entire class of communication tools that go beyond what we're used to in the physical world. For example, techniques like the "attention funnel" can be implemented in-world so that when one user wishes to point out something, the other explorer sees efficient full field-of-view guides to quickly center his attention on the object of interest.

Hands can also be used to manipulate objects, and this manipulation lends itself well to collaborative tasks. It is clear, for example, how if two people grab a large object, its position and orientation can depend on both of their hand movements, in a way that is fairly natural.

Speech is a natural communication affordance that is foundational to a variety of synchronous collaborative activities. It can also be used asynchronously — some examples would be attaching messages to objects for later retrieval or having playback trigger according to interactions or timing. Speech-to-text can also be used as a form of communication. Text can be placed in the peripheral field of view of another explorer to be viewed as attention allows. Drawing tools may also be used as communication tools – by sketching a variety of symbols – words, punctuation, pictographs, and so on. In fact any precisely-timed or rapidly-responsive visual or auditory signal may be used for communication, though improvised communication is likely more useful for play than productivity.

Capture   These tools are aimed at capture activities including annotation, note-taking, and recording. In Section 2.2.5, I presented background research on note-taking, surveying the mechanisms that verbatim and generative/synthesized notes leverage to facilitate learning. Verbatim notes allow the learner to repeat information later, while generative notes deepen learning at the expense of completeness. This tradeoff only exists, however, when verbatim notes require attention and effort to capture. The "recording" affordances for capturing images, videos, audio recordings, and scene recordings provide tools for creating notes that are both verbatim and generative, with minimal effort. Some mechanisms leveraged by such notes are as follows:

- *The Transient Information Effect.* By giving explorers the assurance that experiences can be reviewed and revisited, the systems frees their working memory from the task of verbatim capture for storage in memory. This allows them to get the most out of the experience the

45

first time without sacrificing their maximum yield/absorption over time.

- *Bookmarking.* Combining verbatim recordings with bookmarking yields the advantages of a complete reference together with the benefits of intentional capture. The complete reference guarantees that information is not lost and can be reviewed later, while intentional capture aligns moments of personal significance (e.g. a realization) directly with recordings. A quick, simple gesture allows bookmarks to be made without incurring a significant working memory cost. Prototypes of bookmarking tools will be presented in Section 4.1.3.

- *Memory Cues.* Intentional capture of still images or video clips creates memory cues that can be used during review to reconstruct insightful moments and perspectives from a learning experience.

Review   These tools are aimed at engaging with captured information about prior experiences. The categories of *review, repeat, reflect, recall, and remember* each highlight different kinds of affordances that are useful for the self-regulated learner with regard to this information and the corresponding processes of engagement.

The "Five R's" are review, reflect, repeat, recall, remember. Each refers to a different aspect or attribute of the deliberate learning process. The Five R's need not be separate from the Three Ex's, but engaging with them is a qualitatively different process. One example might be having a look at a snapshot of a previous activity while performing, say, an experiment. This process of review and reflection is made possible through the use of verbatim capture tools, and the availability of affordances for retrieving them at will.

Review. Suggests passive "viewing again." Simplest form of retrieving saved representations.

Reflect. Suggests an active component during review – for example, creating new annotations or writing new notes in the process.

Repeat. In this case, I refer to a form of active repeating of an activity. Examples would be: repeating a saved level of a game (which may be unique due to typical randomness in initialization), practicing a particular set of steps or transition in the middle of a longer sequence (anything from algebra to car racing). This requires that the original "physics" of the world be accessible (more on

this in the section on Simulation).

Recall. In this case, I am referring to affordances that support *practicing* recall. A simple example from the physical world is the use of flashcards – these represent recorded information in a particular way that supports recall-oriented learning practice.

Remember. This is the basic "learning goal" – to *remember* what was learned. It does not have direct significance for tools .

Modality

I refer to *modality* as the physical basis by which digital input is provided to the virtual reality system. Some input modalities are three degree-of-freedom spatial input, six degree-of-freedom spatial input, and speech input. Six degree-of-freedom spatial input devices are assumed as basic equipment for the reference experience instantiations presented in this thesis. Even so, only some kinds of tools utilize six degrees of freedom in an essential way, whereas others require less dimensions, bandwidth, or precision. These different kinds of tools are distinguished as *low bandwidth spatial* versus *high bandwidth spatial* modalities.

High bandwidth spatial. Examples include sketching, hand-written text input, pointing out part of a scene from a particular angle, hand gesturing, nuanced head gesturing.

Low bandwidth spatial. This form of spatial input uses position and/or orientation, but for tasks which require less bits of information. Some examples would be: selecting an orientation, selecting items from a panel, array, or menu, and performing simple symbolic gestures. One reason to distinguish the low bandwidth vs high bandwidth spatial modalities is that the former may be more readily translated to non-6DoF VR and non-VR devices. Although this is not relevant to the core VR experience, it will be relevant to the discussion of *Reality Integration* in Section 3.3.

Speech Input. Speech input can be used in a variety of ways — several of them were recounted above in the section on tools used for the purpose of collaboration (Section 3.1.1). Other examples might be to transcribe text in the environment as part of a process of ideation, creativity, or note-taking. Under some circumstances, speech input may be used for tool retrieval – although in a variety of settings, especially collaborative ones, using speech can be disruptive or confus-

ing. It is also prudent to distinguish *synchronous* from *asynchronous* uses of speech. I apply a rather strict definition of *asynchronous* where speech-to-text is asynchronous because it typically requires waiting for short clauses to be spoken completely before the result can be used. Live or conversational speech interfaces are the only ones that are considered synchronous. Any variant of tagging or voice messaging is accordingly asynchronous.

Scope

The general or specific applicability of each tool is another important input to the design of a spatial toolbox layout. I propose two attributes that are relevant here:

Frequency of use. Some capabilities are used more frequently than others and across many different contexts. Teleportation and other forms of locomotion, for example, are frequently used almost regardless of the application.

Context specificity. Some capabilities are highly relevant to a particular context, while others are more general. Speech-to-text transcription, for example, is more context-specific than locomotion, although in those use cases where it is applicable, it might be used much more than locomotion. On the extreme end, there may be many tools that only exist within a specific context. A Geiger counter, for example, is specific to contexts that explicitly represent radioactivity – which would apply to a narrow segment of entertainment, education, and training use cases.

3.1.2   Tool Ergonomics

In the previous subsection, I introduced a taxonomy for tools. This taxonomy covered aspects of tools that follow directly from their abstract function and purpose . These are independent of any particular instance of that tool in an applied context. This subsection is about *tool ergonomics.* That is, taking into account what we know about the purpose of the tool, the fixed aspects of how it works (modality), and how frequently it will be used in one instance and in general, how exactly will the user access and operate the tool?

Some design guidelines (derived from common user interface design guidelines):

 • *Consistent access for efficiency.* A frequent-use, context-independent tool should always be

found in the same place (e.g. pen, camera, hand), since the user can then leverage muscle memory, and not have to unlearn or constantly think about how the access method may have changed depending on the context.

- *Shortcuts can provide consistent access to changing tools.* Most contexts offer some context-specific tools – therefore it makes sense to consider how to provide efficient access to the subset of tools that do change frequently. In a given setting, there are specific, constrained sets of "quick-assess" locations. One example would be the buttons on handheld controllers. There are typically a small number, and the functions assigned to these will be the quickest to access (not to preclude changing the function of buttons based on spatial location). To reiterate, if all quick-assess locations are populated by context independent tools, then all context specific tools will be slow(er) to access. Hence it is advisable to reserve a quick-access location for a rotating set of context-specific tools.

3.1.3 Examples of Tools

In this section I provide a few examples of tools that fulfill multiple purposes and have a broad scope.

Word and symbol generator Interacting with words and symbols is essential to learning: reading and writing text, drawing diagrams, solving equations, and so on. An empowered learner needs to have consistent, rapid access to these affordances across all learning environments. Speech-to-text is a promising modality for words and some symbols that are easily pronounced. To achieve efficient input of equations and manipulation of existing expressions, a hybrid tool that uses speech and low-bandwidth spatial input might be appropriate.

2D Camera and Image Browser A 2D camera and corresponding image browser allows users to capture 2D perspectives, and a corresponding image browser allows them to view an inventory of their captured images. A 2D camera has numerous advantages, despite ostensibly being non-native to VR. Some advantages are that capture and viewing of 2D images is not resource intensive, and 2D snapshots can be viewed on 2D devices like desktop computers and mobile phones, or even printed on paper. Therefore, I would argue that a 2D camera is indispensable to most learning applications.

Hands A hand tool is multi-purpose and adheres to many natural interaction metaphors. A hand can be used to grab and reposition objects in the environment, or it can be used to gesture for non-verbal communication in multi-user settings.

Locomotion A locomotion tool or affordance is critical to interacting with virtual environments that are larger than the physical VR space. One common method of locomotion is teleportation, where a user points at a desired location with a hand-based high-bandwidth spatial tool, and can be instantly moved to that location. Other methods are possible, but many can cause minor motion sickness.

3.2 Why VR? An Analytical Survey of Learning and Discovery Technologies

In this dissertation, I make a case for using immersive virtual and augmented reality technologies for learning and discovery, assuming a self-regulated, constructivist learning setting. The proposition of using a new technology in learning and discovery settings entails significant development and training costs in many fields. For this reason it is important to be clear about why one might consider incurring these costs – the mere fact that VR supports constructivist learning processes does not on its own justify using it preferentially over other technologies that also support these processes. Using the tool framework presented in the previous section, I analyze the offerings of existing alternative technologies, comparing them to VR. The first section looks at communication and exposure affordances, and the second looks at capture and review affordances. The technologies considered in each case are overlapping but not identical, since some are applicable to one set of purposes or the other. Finally, I draw overall conclusions about the promise of virtual reality based on these considerations.

For the purposes of this analysis, there are two broad categories of activity involved in (optionally social) learning and discovery: the world-oriented activities of exploration, expression, communication, collaboration, and so on, and the self- or inward-oriented activities of capture and review. The ultimate proposition is that immersive, social virtual reality systems can do it all, but as we will see in these two sections, the list of competitors I consider in these areas is different.

3.2.1 Exploration, Communication and Collaboration Affordances

This set of outward or world-oriented activities encompasses information-seeking, exploration, expression, creation, communication, and more. I contend that this "outward" orientation also delineates a bounded and coherent set of current technologies which support these activities.

Technological Media to Compare

Same Time, Same Place Face-to-Face   I refer to settings where collaborators are in the same place at the same time, and communicate face-to-face as an integral part of their activities. Other objects or devices may be involved in this interaction (a paint canvas, a smart board, physical props or models, etc.), but there remains a strong core of face-to-face interaction.

Videoconferencing   Communicating over a video channel (audio and video) is a common practice globally at present, and is provided by numerous commercial services. While a subset of these services support companion interactions such as screen sharing or sharing a drawing canvas, this is rare. Perhaps the most common among these activities would be using a collaborative text editor while communicating over video – however, neither gaze nor direct hand gestures are supported in the shared space as part of the attention coordination process.

Collaborative Text Editing   One of the most common modes of remote collaboration is collaborative text editing. Two or more users can simultaneously edit a linear 2D document involving text, images, and other embedded media. Audio or video conferencing technologies may be used in parallel to enhance coordination and collaboration during the process.

Non-Immersive Virtual Environments   Online games and social environments are the most common place to find communication and exploration in non-immersive virtual environments. The "MMORPG" category of game, along with the renowned "Second Life," which sought to be a social platform for general purpose social communication and entertainment. While these games (accompanied by real-time audio communication) do give users an awareness and interaction with other users, the avatar representations are not mapped to physical body movements, but are rather controlled by indirect manipulation through keypresses, joystick turns, and so on.

Immersive Virtual Environments These environments are my primary object of study in this thesis. Multiple users share a virtual environment in which their body movements are mapped directly (in position and orientation) to their avatars' apparent body movements in the shared space.

Affordances and Support in Varying Media

In this section, I consider affordances for exploration, communication, and collaboration across the various media introduced above.

- **Point.** This refers to deliberately using the hand or other means to indicate a point of focus in a shared space.

- **Gaze.** Similar to pointing, gaze indicates a point of attention, but does so continuously and incidentally (rather than deliberately).

- **Manipulate.** Creative or constructive collaboration requires manipulating objects or representations related to objects in a shared space. This can be done "directly" in space, or indirectly using other input methods. Some media allow this to be done natural in a simultaneous way, while others require one user at a time to control or modify objects.

Point The idea of pointing comes from face-to-face communication. That is, the act of one person orients her/his arm in such a way, in a shared physical space, that another person can determine or estimate a remote object or location of interest. As such, it is well-supported in face-to-face settings. In videoconferencing, pointing is essentially not supported – in the typical setting, there is no shared space – but rather two one-way transmissions that happen simultaneously. The closest thing to pointing during a videoconference would be using the mouse to indicate an area of interest, although typically only one of the collaborating parties would be able to do this. There are commercial products that aim to improve on some of these details, but overall the medium is not oriented towards this kind of interaction. Collaborative text editing does support pointing quite well, since users' can typically see other users' cursors and highlights. Annotations enhance the feature even more, by allowing "pointing" to become persistent. In a non-immersive virtual environment, it is common for a means of pointing to be included – using some form of crosshairs

or other indicator projected onto the shared visual space. In an immersive virtual environment, pointing works analogously to face-to-face, but can be easily augmented with visual or auditory support to make it easier to find the object of attention, or increase the precision with which it can be specified.

Gaze   In same-time, same-place communication, gaze is used as a continuous indication of attention – looking at the speaker, is different from looking at an object in the environment. Peripheral aspects of gaze such as head movement and facial expressions further increase the expressive power of this communication affordance. In videoconferencing, nodding and looking towards or away from the camera carry some expressive power, but subtle inferences about attention made from eye movement pale in comparison to what is found in a shared physical space. As noted previously, there is no shared space in a videoconference, so gaze can scarcely indicate something in a shared space. When considering collaborative text editing, pointing and gaze are replaced by the same mechanism, which is the cursor location. This does continuously indicate the focus of attention when the user is actively engaged, but fails to convey where they are actually gazing with their eyes. In a non-immersive virtual environment, any anthropomorphic avatar (which would usually entail having a head) possesses a notion of gaze. Because head movement is controlled indirectly, its gestures are deliberate. Head gestures like nodding can be activated, but looking towards or away from the speaker to indicate attention is artificial at best, and most likely non-existent in practice. In an immersive virtual environment, head gestures translate very well, since the visual rendering system, which needs a precise measurement of head movement to show the viewer the proper perspective, can use this data to render that head movement for other viewers. What is sometimes referred to as *head gaze* indicates visual attention through head movement. For the moment, measurements of eye movement in VR are rare, but this appears to be poised to change imminently.

Manipulate   Manipulation refers to changing the environment deliberately. In the physical world, these methods are well-known – grab, carry, push, pull, etc. The collaborative dynamics are intuitively understood as well – collaboratively manipulating a small object may not be possible, while manipulating a large object may not be possible alone. In a videoconference, participants can manipulate objects in their own environment – pointing the camera at a whiteboard, for example,

	point	manip	direct	gaze	simul
same time, same place f2f					
video conf					
collab text edit					
non-immersive virtual env					
immersive virtual env					

LEGEND – Affordances

point	pointing
manip	manipulation
direct	direct pointing and manipulation
gaze	indicating attention with gaze
simul	simultaneous manipulation

Table 3.1: Comparing Media's Exploration and Collaboration Affordances

allows them to use manipulation (writing on the board) as a means of communication. Simultaneous manipulation isn't possible there because there is no shared space – objects are either in one participants' environment or the other. In collaborative text editing, there is indeed a shared space, and both users can manipulate it at the same time. It deserves to be noted that the seamless experience of collaborative text editing which is taken for granted by users is in fact an impressive technological feat (its impressiveness is not as readily apparent as, say, wireless communication or a VR display). In non-immersive virtual environments, collaborative manipulation is sometimes supported, albeit in an awkward way. Manipulation is in essence a continuous physical motion, whereas manipulations mapped through other input devices are lower-bandwidth and less fluid. In an immersive virtual environment, manipulation is well supported. Although today's handheld controllers may trigger "grabbing" with a button press (unnatural), once an object is grabbed, the ensuing smooth, six degree-of-freedom controller of the object's position and orientation is high fidelity and natural. Collaborative manipulation may even offer advantages over face-to-face – it is easy to implement a simple policy in software that determines who wins when two people want to manipulate the same object (in a non-collaborative way) at the same time (e.g. last to grab). In the physical world, the tug-of-war mechanism has many shortcomings (object breakage, conflict escalation, etc).

3.2.2 Capture and Review Affordances

This section discusses affordances relevant to capture and review, and investigates how each does or does not manifest itself in existing technologies for learning and discovery. Table 3.2 summarizes what is discussed. Not every table cell is discussed explicitly in the text, but the discussion of those that are should allow the user to extrapolate what is meant.

	comm	cont ctrl	pace ctrl	rep inst	rep delay	verb pers	verb	notes
print		rigid						
video – recorded		rigid	rigid			poss		
one-on-one tutoring		soc	soc		soc			
videoconference	rigid	soc	soc		soc	poss	poss	
classroom F2F	depends				depends			
VR – interactive				tough	poss	poss	poss	poss
VR – interactive + live		soc	soc	tough	poss	poss	poss	poss

LEGEND – Affordances

comm	real-time communication/interaction
cont ctrl	content control
pace ctrl	pace control
rep inst	repeat at will – instant
rep delay	repeat at will – with brief delay
verb	verbatim capture
verb pers	personal experience verbatim capture
notes	note-taking

LEGEND – Support Modifiers

soc	subject to social factors
poss	possible to support
tough	future – difficult
depends	depends class size
rigid	inflexible

Table 3.2: Comparing Media's Cap-Re Affordances

Technological Media to Compare

Print   The print medium refers to any form of written text. For the purposes of this analysis, I do not make a distinction between text on a digital display versus text on paper. There are clear and important differences between these (e.g. search functionality), but they are not salient to this analysis.

Recorded Video   Here I refer to linear, pre-recorded camera footage or visual animations produced by any means. For the purposes of this analysis, there is no distinction between 2D video, 3D, 360, or 3D/360 video.

One-on-One Tutoring   This is frequently assumed to be the gold standard of teaching – individual attention is given to one student, learning materials are frequently pre-prepared, and the tutor has the ability to alter the style or pace of presentation according to the needs of the student.

Videoconference   I refer to one-on-one videoconferencing, which enjoys many, but not all, of the advantages of face-to-face tutoring. This will be revisited in the section below.

Classroom, Face-to-Face   Face-to-face classroom instruction is in some ways similar to one-on-one tutoring, but the presence of more students decreases the amount of attention and cus-

tomization for each individual student.

VR – Interactive An interactive virtual reality application puts the user in control of exploration, within the constraints some particular environment and subject matter. Other kinds of media may be embedded, but are assumed to be used in short segments (else, for example, the experience would fall into the category of *print* or *recorded video*, despite being presented in a VR headset.

VR - Interactive + Live In this medium, multiple explorers occupy the same world at the same time. In general, my assumption is that users maintain their autonomy, with working together made available optionally. In practice, however, this is a design choice made by an experience developer.

Affordances and Support in Varying Media

Real-time Interaction The real-time interaction media are tutoring, videoconference, classroom instruction, and live VR. Interacting over videoconference is real-time, but it is not face-to-face, and therefore has some shortcomings and limitations. Receiving instruction in the classroom may be more or less interactive depending on class size, pedagogical model, and so on. Using a baseline 6 DoF VR system for real-time interaction has some advantages and some disadvantages as compared to face-to-face, and is considered a well-supporting medium.

Autonomy This term refers to the degree and type of control the learner has over the content being presented. I distinguish between *pace* autonomy, where the learner is in control of when and how fast content is presented, but not the content itself, and *content* autonomy, where the learner decides in some fashion what she wants to learn.

Repeat at Will This refers to the ability for the learner to repeat the presentation of a particular segment of content at any point in a learning session. I distinguish between *instant* and *with brief delay* variants of this feature. When learning from **printed materials** (text and images), the learner always has the ability to re-read a sentence, glance again at a diagram, and so on; that is, the ability to repeat at will instantly. In contrast, when **watching a video** individually (e.g. on a laptop), the learning can repeat a segment, but the process is slower – whereas text provides a continuous visual index of what has been read or seen, a video must be browsed by jumping, and

after jumping, the learner requires at least a second or so to determine whether the right location has been found, and if not, whether the sought location is before or after the erroneous current location. While it is possible to imagine (and research in this area has demonstrated) faster ways of browsing video, they are uncommon and difficult to realize. So for my purposes here, reviewing video will be considered slower than reviewing text. In the case of **one-on-one tutoring**, instant review is not possible, but by interrupting the tutor, the learner can achieve the review of a concept quickly, albeit somewhat differently than with video. The effectiveness of this will depend on the individual tutor, and how accommodating he or she is of questions.

Verbatim Capture   I use this term to refer to a technology-enabled capture of the content that has been presented that is, in a certain sense, "identical" to the original when it is replayed. **Printed material**, for example, by its nature can be repeated and reviewed (as highlighted above). In that sense, its content has already been captured. I make the additional distinction of verbatim capture of the *personal experience*. With the example of printed/written text, two versions of capturing the personal reading experience would be (i) an eye-tracking video, which shows in sequence which words a learner gazed upon, or (ii) a screen recording which shows how a learner scrolled through text written on screen. These examples are somewhat esoteric, so I consider this not to be a feature of text as such. Viewing **recorded videos** is analogous to text as regards verbatim capture. Next, considering **one-on-one tutoring**, it is possible to use a video camera to record the contents of a session. Audio recording may be the most common method, although even this is uncommon. A video camera may also be used, but capturing a complete set of relevant visual angles (desk over the shoulder, tutor's face, any screen that may be used for digital content) in a clearly viewable way is typically not feasible. As such, I'll consider this not to be a feature of one-on-one tutoring.

Note-Taking   There are two categories of note-taking that are distinguished in the related literature – *verbatim* and *synthesized* notes (see Section 2.2.5). As defined above, hand-written notes are not a form of verbatim capture – compared with an exact transcription or an audio recording, hand-written notes, even those that are written without much intermediate thought, are not complete in the same way. The distinction in the note-taking literature concerns the difference in cognitive processes behind the two kinds of handwritten or typed notes, but in this case I am

57

concerned only in the quality of the reproduced experience. Handwritten or typed notes do not qualify from this perspective. Instead, I refer to the entire genre of annotations that are active and specific in time or space as note-taking activities. This includes bookmarking a time, assigning a keyword, drawing a sketch, underlining, recording a voice note, and so on. All of the **non-VR media** allow learners to take notes in different ways. As a baseline, they can bring a notebook and take handwritten notes. In the case of **VR** (with or without live interaction), note-taking features may be built in a straightforward way, but are not an inherent feature of the systems themselves. These might include spatial or temporal bookmarks, 2D or 3D snapshots, voice notes, sketches, and more.

3.2.3 The Promise of Virtual Reality

In the previous sections, I established some ways of thinking about what existing learning technologies "do" in terms of supporting optionally-social processes of learning and discovery. In this section, I will apply that thinking, and make some additional claims, about the promise of virtual reality – or, more generally, *virtuality* – for these purposes. My claims about the advantages of virtual reality fall into the following categories: (i) there are some additional fundamental advantages of VR in social and abstract cognition, (ii) it does what alternatives do, but better, and (iii) while alternatives only cover a subset of affordances and needs, you can do almost everything at the same time and place in virtual reality.

Unique Advantages of VR

I begin with a discussion of the unique advantages of VR. These are the places where VR is fundamentally more powerful, and the alternatives aren't directly comparable.

Go places you can't go physically. Virtual reality allows learners to immerse themselves in locations that aren't physically accessible. Some aren't accessible because it would be too expensive or is too limited as a resource– like the international space station, while others just wouldn't be physically possible – like standing inside of a human cell.

Do physically impossible things to your environment. Unlike reality, in virtual reality you can rewind, like a video, or copy-paste, like in a text document or image editor. Only it's the whole

58

world instead of just a digital artifact on a screen. Many more such possibilities will be mentioned in later sections and chapters.

Modify social reality. In the face-to-face world, there are certain aspects of reality we can't change – like gender and race, and those that are hard to change, like clothing as a status symbol. These attributes influence how we interact with each other, in a way that can perpetuate entrenched prejudices and divisions, and can interfere with evaluating or engaging with ideas on their own, as opposed to in connection with the people proposing them. In virtual reality, the physical appearance can be altered at will. Race or gender can be obscured or changed, along with attributes like physical height. While this is all possible, for example, in non-immersive virtual environments as well, the reason it is so significant in immersive virtual environments is that so many other aspects of face-to-face communication are preserved. Head gestures, hand gestures, and so on, along with group dynamics, like taking turns, looking at the speaker, etc, are all transferred seamlessly into VR.

Outperforming Alternatives Head-to-Head

In the discussion of affordances offered by different learning technologies, I touched on some of the advantages offered by VR in each case. In this section, I bring these points together in one place in an attempt to form a coherent picture. Let us consider first the affordances for collaboration and exploration.

Communication and Exploration As shown in Table 3.1, Immersive Virtual Environments are the only medium that supports all of the examined affordances to the maximum degree. For pointing, immersive environments offer the clear advantage that the object of interest can be highlighted directly in the environment, or even directly in the field of view of collaborating parties. The same applies to communicating gaze in virtual reality. Addressing manipulating, objects can be manipulated in more sophisticated ways than physical objects (e.g. any object may have its shape or size changed), and if multiple users wish to modify the same object, copies can be made so that all can have their way.

Capture and Review The advantages of virtual reality in the area of capture and review are profound. Forms of capture such as audio or video capture are inherently limited to a single per-

spective. Recordings in virtual reality allow learners to see different perspectives upon review, and which has vast implications for the depth of understanding that a single experience can yield. It can be hypothesized that virtual reality recordings are more powerful memory cues than their audio or video alternatives, since more aspects of the environment are reproduced (wider field of view, objects that are completely out of field of view, but of which user is aware, and so on). In the case that learners are interacting with simulations, recorded materials remain interactive, so that learners can explore one and many "alternative futures" – answering the question "what would have happened if?"

If we look now at the affordances that were highlighted in the previous sections (Table 3.2), it can be seen that many of the "VR – interactive + live" affordances are highlighted in yellow. This means that these features do not come "by default," but can be built. As such, this ties directly into the discussion in Chapter 3.3 of what must be done in terms of design and technology to build a true *medium* of virtual reality (or *virtuality*). For example, it is clearly *possible* to build a live instant replay feature in VR – it would behave just as watching a video on a website. At any point, one can jump back with no delay, and see a segment again. In virtual reality, the same could be done with the entire immersive experience, but of course this feature will not exist in applications or platforms unless it is actively constructed by developers and technologists.

All Affordances in One Place

The final point I'll make about the promise of virtual reality is that it offers all of the affordances discussed, in the areas of exploration, collaboration, capture, and review, in one single environment where all are interconnected. In typical learning environments, the use of different media, although complementary, is also disjointed. For example, a learner may take notes on paper while viewing a video. While certain notes may be relevant to certain scenes, there is generally no clear way to refer to the scene (handwriting timestamps might be the best), or to navigate to it if it is identified. As another example, the use of collaborative document editing is very powerful within the constraints of linear, 2D documents, but the use of non-verbal communication between authors is only possible if they happen to be in the same space. Even then, authors need to look away from the document, communicate, and then look back at the document– another disjointed process. VR offers nearly

infinite flexibility to link and integrate these different aspects of the exploratory learning experience into one place. Referring to the two examples just mentioned, handwritten notes in VR can be automatically linked to videos they refer to, improving the review process, and shared with others, to benefit their capture and review process as well. Collaborative document editing can be brought into a shared virtual space where users can gesture at each other while discussing and pointing at the content of the document. Add to this the ability to step into the video instead of just watching it in 2D, or the ability to collaboratively edit the said document in a virtual atelier with relevant reference materials collaboratively arranged in space, and the advantages of VR become even more vivid.

3.3   Reality Integration: Towards Virtuality as a Medium

Through the explorations described in Chapter 4, it became clear that the needs for the explorer's tools are not limited to the VR experience. For example, the ability to review learnings cannot be constrained to the HMD and the ability to communicate and collaborate cannot be limited to others in VR. These realizations pointed to the need for a broader conception for how virtuality can function as a component in an ecosystem of learning.

In this section, I will consider the problem in an abstract way, and build a historic analogy to the growth of an automative infrastructure out of the core innovation of the combustion engine. Then I will highlight specific shortcomings of the consumer hardware offerings and ecosystem today, in a way that is general and applies equally to the learning space. Finally, I will come back to the Equipped Explorer and introduce the notion of *Reality Integration* that is required to make virtuality function as one part in a heterogeneous whole that facilitates learning.

3.3.1   Today's VR is Not a Medium

The term "medium" is extremely important in understanding information technology and its effect on people and society. The idea championed by Marshall Mcluhan in the 1960's was that people, as receivers of messages conveyed through a medium, easily become unaware of the presence of the medium, although it has a profound impact on how they evaluate that message [MF67]. In addition to this property of being immersive (and coercive), I would also highlight another critical

61

aspect of a medium: it supports distribution, modification, and redistribution in a way that is self-apparent and self-explanatory. For the rest of this section, I'll refer to this as the *self-explanatory sharing* property. This is where today's VR is deficient, and I will take this opportunity to draw an analogies to the print medium and make more explicit the kind of technological development that is needed.

The properties and affordances of the printed page are self-apparent and self-explanatory. You can hang it, put it in your briefcase, make a photocopy, cut it with a pair of scissors, and so forth. Not only that, but you can give it to your friend or colleague, who can equally well make a photocopy, cut it with a pair of scissors and hang it on the wall. VR has none of these properties: it isn't clear what you can do with it, and when or how you can give it to someone. There are many questions users ask: Can you see what I'm seeing? Can this experience be shared? In-person? Remotely? Real-time? After-the-fact? Can I record it? Or for the savvy: if I have an experience in VR and I record it, can I look around at the other parts of the scene that I didn't see the first time? The answer to all of these questions at present is "It depends," with many qualifications that depend on the particular hardware device and software application.

I argue that the self-explanatory sharing property is important for both individual and social purposes. The individual, if she is going to be exposed to learning content using VR, must be able to understand the exact ways in which the content will be available for review. The transient information effect from cognitive load theory shows directly the impact that this awareness and assurance has on the ability to learn. Contemplating *will I be able to refer back to this scene while I'm on the bus home after class?* is not something that should be occupying the learner's mind during a learning experience. VR learning platforms will be handicapped until the "content" is able to move in a seamless and self-apparent fashion from inside to outside and back again.

The social utility of the self-explanatory sharing property lies in the assurance that peers and experts can be consulted and collaborated with no matter what. Learners must not end up in a situation where a question cannot be posed or the answer cannot be offered because the relevant person doesn't have a VR device. There are unavoidable constraints on how this sharing can happen, but their properties are self-apparent and self-explanatory. For example, if I use 2D snapshots

to share aspects of a scene with someone by email, I know that I must photograph all the relevant angles manually. It is less comprehensive than sharing a 3D model, but it is clear what needs to be done. Today there is no expectation that every platform or application provides the VR user with a camera tool, so VR users cannot depend on even this simple form of experience sharing.

At this point I've argued that VR lacks certain properties of a medium, and explained why these are important in the context of learning. Next I'd like to provide an analogy that sheds light on the magnitude and nature of the task of making VR (or *virtuality*) a medium.

3.3.2 Historical Analogy: From Combustion Engine to Ubiquitous Automotive Infrastructure

In order to clarify what it might mean to build an infrastructure to make VR a medium– at a minimum giving it the self-explanatory sharing property– I offer an analogy to the history and functioning of the combustion engine. What is exciting about the combustion engine, in hindsight, is that enabled ubiquitous individual rapid transport. Today, this shapes everything from our urban landscapes to our psychological sense of freedom. But how did this come to be? The combustion engine created the possibility of a vehicle – the rotation of a crank or a shaft clearly makes it possible, through the use of wheels, to create linear motion; and the use of axles then provides a set of support points upon which it is possible to build a frame that can be used to transport objects. It is clear that although the combustion engine is a key enabling component of a vehicle, it is not a vehicle in and of itself. In the same way, I would argue that the 6 DoF tracked virtual reality system provides a core capability that is necessary for making virtuality into a medium, but it is not sufficient on its own. To take the engine analogy a step further, consider how the automobile's utility relates to the infrastructural support it enjoys in industrialized societies. Without paved roads, automobiles are limited in speed and usable area. Without plentiful gas stations, usability is similarly impaired. As such, the combustion engine required two layers of system integration to reach the full potential that it has achieved today (autonomous vehicles promise to take it to yet another level in the coming years). I argue that we can declare VR a medium at the point where its boundaries vis-a-vis the real world and the rest of the virtual world can be described as seamless and self-apparent transitions. Only then will it realize its potential to augment our experience of the world we know, allowing us

to better learn and collaborate with others. Beyond that initial accomplishment, there will be other levels of further integration that take its utility to new levels.

3.3.3   Shortcomings of the Present Consumer VR Technology Offering

I will elaborate first on what I propose are the "vehicle" requirements for VR, and then speculate on further levels of ecological utility that are made possible by different kinds of infrastructure. I propose four kinds of "contextual integration" that are required of a "reality medium": temporal, spatial, social, and informational. First, I elaborate on the shortcomings of the virtual reality headset as such. Some of these shortcomings are well recognized, while others are less so.

The User Is Walled-In   The fact that the virtual reality headset separates him or herself from the surrounding physical environment is at one level obviously a natural consequence of its core functionality. On the other hand it represents a serious practical **safety** problem – most environments aren't empty, some change unpredictably (imagine a cat pushes a skateboard into a VR space), and so on. The state of VR is unfinished and unacceptable until this problem is consistently and fully solved.

Bystanders Walled-Out   Unlike watching a movie, reading a book, dancing, or anything else one can do in a space with others, the VR headset as such creates an immersive experience that is invisible to others. While there are undoubtedly circumstances where this is desirable to the user, there are many (perhaps even more) where visibility into the virtual world would be highly desirable and advantageous. The most common solution out there right now is to share the first person point of view on a screen. This barely begins to approximate the concept of a transparent or a shared space. The screen viewer can't look anywhere the VR user isn't looking in the space, and even those places she is look at are hard to focus on because the head moves so frequently and unpredictably.

No Consistent Identity   To the degree that people have been experimenting with social VR in the past two years, it remains the case that there is rarely a consistent notion of visual identity between "apps" or "games." So long as this is the case, each application represents a virtual silo that is completely independent of all others. With this mode of operation, it will not be possible to

conceive of virtuality in the same way as reality – a place where you go many places and "where many hats," but underneath the costumes represent one consistent psychological human identity. At present, the aspect of your identity you're most likely to be able to bring with you is your voice. This is clearly a start, but there's a long way to go.

No Consistent User Interface or Operating System　It is true that designing interfaces in virtual reality is challenging. It is still being explored, best practices have yet to be established, and there are few conventions to go on. Even so, to reiterate and build on the previous point – in order to fluidly transition between many different virtual environments, users will need consistent aspects of their user interface (what I referred to above as a *toolbox layout*). If we consider the SteamVR platform today, the consistent user interface is the physical "system menu" button, and software system menu that then appears. This represents a hard boundary beyond which no other consistency is allowed to go without individual developers independently adopting similar conventions. If we imagine explorers spending many hours of their days inside some form of virtuality, we cannot at the same time erect blunt barriers that limit the flow of experience between "apps" that each get to choose our avatar and toolbox for us.

3.3.4　Facets of Reality Integration

In the previous sections, I have argued that today's VR is not a medium as such. I highlighted specific shortcomings in a general context. But what does this mean for the Equipped Explorer? In this section I will introduce the concept of *Reality Integration*, framing each of its parts in terms of needs for learning.

Reality Integration refers to making VR experiences functional in contexts and goals that are not anchored to VR technology itself. For example, if I wish to learn algebra, VR may be used as a tool, but the goal is the learning, not the use of the tool. Similarly, if I wish to collaborate with peers on the design of an autonomous drone, VR may help us to create the 3D components, and debate matters of aesthetics, but the point is to build the drone. In this dissertation, I argue that VR is highly applicable to learning; but I will take this opportunity to affirm that the goal is always for learning to take place, not for VR to be used. This point became highly salient through the explorations described in Chapter 4, in that the question of *integrating* with the non-VR world always became a

central aspect to making VR a useful, viable tool.
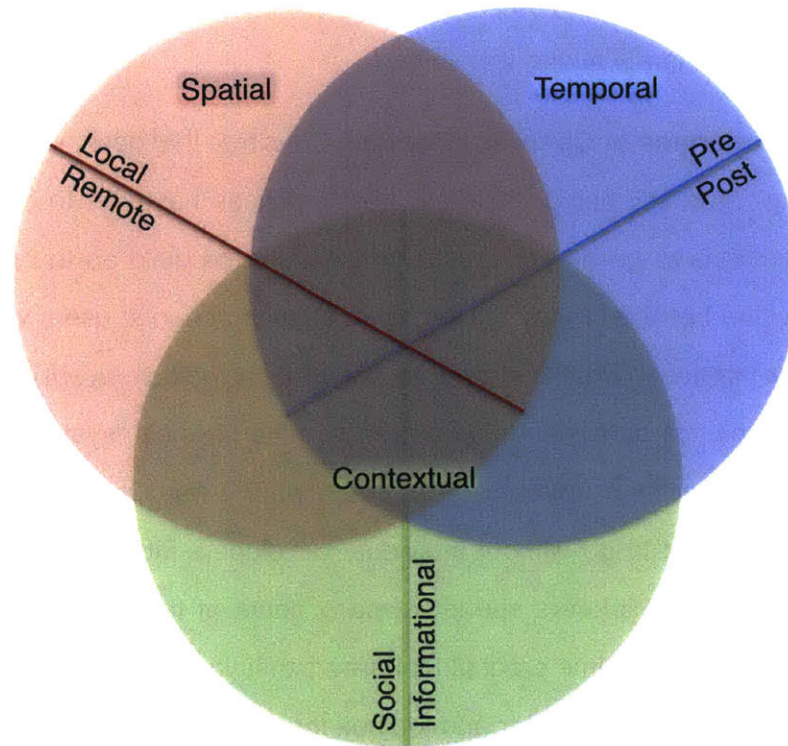
Figure 3-4: Virtuality as a Medium: Forms of Reality Integration

I propose breaking down Reality Integration into three broad categories: *spatial*, *temporal*, and *contextual*, as shown in Figure 3-4. There are certainly other possible categorizations, but this provides a useful starting point. I will elaborate what is meant by each below.

Temporal Integration

There is life before and after each episode of engagement with virtuality. Before the episode, one might prepare to perform a certain task, or have a certain experience. **Pre-experience integration** means being able to properly plan, set up, and prepare what will be experienced or done in VR. Afterwards, one might like to review what happened, or continue to build on it using another set of tools. **Post-experience integration** means being able to review, share, and use digital artifacts generated during an episode of engagement.

Spatial Integration

This form of integration refers to the appropriate use of physical space, with the general principle that there should be one seamless space whenever possible. **Local** spatial integration refers to the inclusion of objects or people in the physical vicinity of a user of a VR system. When other people are present inside or outside the system, their understanding of what the user is seeing or doing should be represented, at a minimum, in one seamless physical and virtual space. I present an example of this in Section 4.3.3. Today it is common for bystanders to glance back and forth between a user in a headset, and a peripheral flatscreen display showing his/her first-person point of view. This separation of representations makes it difficult to reconcile what is going on, and adds significant cognitive load, whether or not it is successful. **Remote** spatial integration refers to a usable configuration of remote physical spaces in a shared spatial context. Two rooms of the same size, for example, can be easily mapped one-to-one, but there are still caveats. In the case of a rectangular room, the choice of symmetric configuration could affect system usability (e.g. locating bystanders in a convenient location). Second, the principle of *one seamless space* exists in parallel to in-scene locomotion. My use of "avatar shadows", introduced in Section 4.3.1 demonstrates how always-on wireframe avatars use an alignment of local or remote physical spaces to enable in-world users to teleport away, or join back together in a consistent spatial relationship to one another.

Contextual Integration

The two kinds of contextual integration I'll discuss involve people and information, respectively. **Social integration** refers to aspects of an experience that relate to other people. Others can be equally immersed in the space, observing passively, viewing a recording, or seeing a picture afterwards. They may also access aspects of an experience less directly: a group of students may share with one another digital artifacts made in VR without needing to see the process of creation portrayed directly. When people are physically collocated, spatial integration often leads to social integration.

Informational integration refers to appropriate access to information and resources related to the task at hand. An application that portrays terrestrial data should use a GIS database, while a

shopping app should have access to 3D models, pricing, and availability. This is not unique to VR, but VR does pose its own challenges. For example, a web-based shop typically only has 2D assets, while in VR, 2D views will seem deficient. The same would be true in an engineering textbook full of 2D schematic drawings. These are more informative and useful in VR when presented in 3D, and such assets need to be gathered for this purpose. At the intersection of social and contextual integration would be providing remote access to subject matter experts.

Application of the Reality Integration Taxonomy

The purpose of the Reality Integration taxonomy is twofold: (1) to provide a prescription for real-world applications, where the required forms of Reality Integration can be reviewed as part of the design process; and (2) to present a vision of ubiquitous virtuality, where all forms of integration are enabled in all applications. This vision requires the progressive integration of existing technology, and the development of new technologies that aim to perform integration, and are designed with an eye towards modularity and interoperability. Some of the tools presented in the next chapter also provide design exemplars that can guide implementation of specific Reality Integrations.

3.4   Framework Summary

This chapter began by introducing the first part of the design framework: the Equipped Explorer. This establishes learners' needs in VR, in terms of the kinds of activities they perform, and what tools for interaction are needed to support these activities. These activities were divided into the overlapping categories of purpose: exploration, communication/collaboration, capture, and review. Tools are then said to have three defining characteristics: their purpose, modality, and scope. Modality refers to the means of providing input, and scope refers to the breadth and frequency of their applicability. In the next section, I provided a detailed comparison of VR to alternatives for two different categories of purpose: (i) exploration and communication, and (ii) capture and review. Each of these seems to define a natural grouping of alternative methods and media. Finally, I introduced the second part the design framework, which I refer to as *Reality Integration*. This addresses different considerations regarding how VR learning experiences can be integrated with the world outside. I presented a set of categories to provide structure in thinking about these needs while designing a concrete application. Several explorations in the next chapter function as design

exemplars for Reality Integration.

Moving forward into the concrete explorations and experiments presented in the next chapters of this dissertation, the general structure is as follows: the background chapter highlighted certain cognitive and instructional principles that I claimed could be leveraged in VR. The framework section provides a method for describing and prescribing the affordances of systems, from the perspective of the in-VR experience, and its relationship to the world outside VR. In the prototypes chapter, the goal is to sample the four categories of purpose, and in each case leverage cognitive or instructional advantages of VR. In the experiments chapter, I provide empirical evidence of claims related to (1) the learning advantages associated with VR interaction, and (2) the feasibility and utility of communication and collaboration in shared virtual spaces.

Chapter 4

Prototypes and Tools to Equip the Explorer

In the previous chapter, I introduced two major design framework components: the Equipped Explorer, which provides a taxonomy of interactions related to learning in VR, and Reality Integration, which enumerates and categorizes the concerns that arise when incorporating VR learning experiences in real-world contexts. In this chapter, I present a number of prototype experiences, leveraging specific uses cases and content as a vehicle for developing generalizable approaches to facilitating learning in VR. In Section 4.1, I discuss prototypes focused on the single-user experience, looking at what kinds of visualize methods are promising, in what ways the learner can interrogate the environment in the process of exploration, and how the process of exploration can incorporate processes of capture for later review. Then, Section 4.2 focuses on multi-learner experiences, investigating how experiences that work well for single users translate into a social setting. Finally, Section 4.3 is devoted to the development of software and hardware tools that solve problems of Reality Integration and enable applications to be better incorporated into real-world settings.

4.1 Exploration, Capture, and Review

The topic of this section is a series of prototypes that explore the single-user focused activities of the Equipped Explorer: exploration, capture, and review. Section 4.1.1 presents a prototype application for exploring electrostatic phenomena. Then, Section 4.1.2 talks about an application that uses 6DoF spatial input to explore a different kind of system– a neural circuit from the retina. The next two sections both involve systems that support capture and review, first in the context of neuroscience and biology (Section 4.1.3), and then, volumetric terrestrial data (Section 4.1.4).

4.1.1 Electrostatic Playground

This project was a collaboration with Gabriel Fields and Professor John Belcher. Gabe did all the implementation and made significant contributions to the interaction design, and John gave us significant guidance throughout the process, including interaction design, visual representations, and numerical algorithms we were able to port to our VR environment.

The driving vision for Electrostatic Playground was to see and interact in real-time with representations of electrostatic phenomena that are otherwise typically only seen in 2D, and remain difficult to grasp. Using head movement to see multiple perspectives, and 6DoF spatial input with handheld controllers, it should be possible to gain deeper intuition about the relevant phenomena. Some of these are: how electric field lines permeate space, the time dynamics of systems governed by inverse square law forces, and how stable configurations of particles form in space. The project includes affordances related to exploration, capture, and review. In this section, I will first describe the user experience, and then delve into detail on the implementation and design decisions that were made in the course of development. Finally, I will reflect on lessons learned, extracting generalizations and framing them in terms of the Equipped Explorer framework.

Equipped Explorer Affordances

Below I will present the learner experience in terms of the Equipped Explorer activities of exploration, communication, capture, and review. Despite the fact that this project includes collaboration affordances, it is included in the single-user section. This is because the implementation of communication and collaboration affordances was very preliminary. Further exploration and discussion

of communication and collaboration is saved for Section 4.2.

Explore   The explorer has four primary exploration affordances, which are the ability to: (1) create charges of configurable magnitude, (2) grab and throw charges, (3) activate a constant external electric field, and (4) play/pause the simulation. These are illustrated in Figure 4-1.

Communicate   Multiple users can share the virtual space, and the simulation reacts appropriately to charges added or moved by any user. All users have the ability to play or pause the simulation, and the effect of this is seen by all users. Users are represented by simple headset and controller avatars.

Capture and Review   The learner has the ability to record an interaction involving one or more users, and watch a VR replay of the interaction.

Discovery of Affordances and Properties of Electrostatic Charge Systems

There is a spectrum of discoveries that learners are able to make that range from extensions to their interaction capabilities to deep physical insights. Learners may discover that charges collide with other charges and the wall of the space, allowing for billiard-like interactions that move charges from one place to another. Most notably, this allows learners to interact with charges that are out of their direct physical reach (without having the system support remote pointing and manipulation). Next, the external field can be used to accomplish a similar purpose. Charges that are out of reach on the ceiling of the space can be brought down by turning on the constant electric field in the direction appropriate to the sign of the charge.

What might be called the deepest discovery enabled by this application concerns the formation of stable configurations and the dynamics of this type of formation. The rate of reaching equilibrium states varies drastically according to the initial distance of the particles, as dictated by the inverse square law of forces existing in the system. Given the time and the curiosity, learners can form shapes such as a chain, with alternating +/-1 charges, or a tetrahedron with a +4 and four -1 charges.
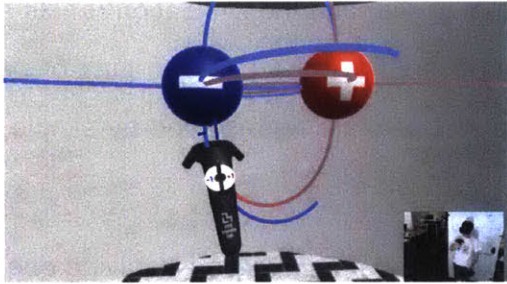
(a) Create charges
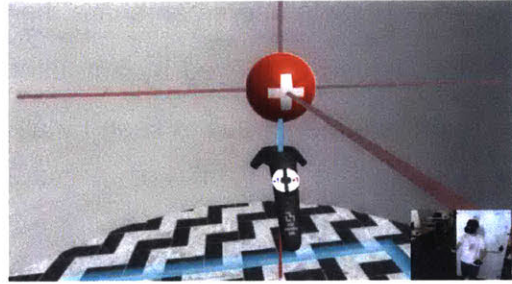
(b) Select charges
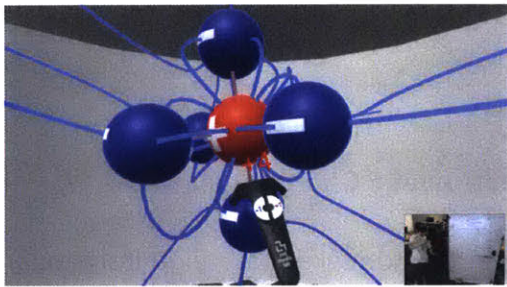
(c) Charges of different magnitudes
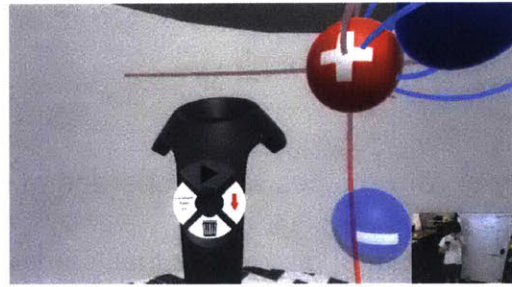
(d) Play/pause simulation
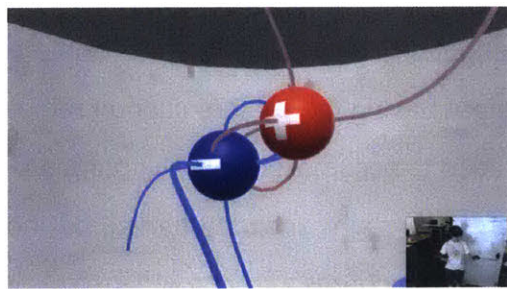
(e) Activate external electric field

Figure 4-1: Electrostatic Playground Interaction Design

Design Choices Related to Representation

A plethora of design choices were necessary in order to be able to effectively implement the affordances enumerated above. In this section I will discuss choices related to the representation of charges, dissipative system dynamics, electric field lines, and the constant external electric field.

Charges and Dynamics The simulated charge dynamics follow the set of forces enumerated in Table 4.1. The implementation details will be discussed further in the next section, while this section will be devoted to the design decisions entailed by and contained in these equations.

To begin, the perception of charges as particles that can collide and do not pass through one another is created by implementing the "Pauli force" shown in the the table [DBB+03]. The use of the Pauli force and the particular parameter chosen for it is likely to vary from application to application within the space of electrostatic visualization.

The Pauli force allows us to implement collision, enabling the user to explore the space in different ways – throwing one charge at another charge can only exert a force on the other charge if the two can collide. On the other hand, the movement of a charge that passes through another charge still affects its dynamics, but in a different way. This could be appropriate, for example, representing particles that are actually much smaller than the spheres shown.

When using a Pauli force, a parameter is chosen to determines how large the force is. The Pauli force is defined by an inverse polynomial, so that objects that get very close experience a large repulsive force, but the contribution of this term vanishes at further distances. If the force gets too large, it can cause energy not to be conserved, so that particles enter "runaway" oscillations, hitting each other harder and harder until they fly off to infinity. The threshold at which this unstable (and unphysical) behavior occurs also depends on the magnitude of the dissipation term. For this reason, I argue that the Pauli exponent should be empirically determined after the dissipation term has been fixed as described previously.

Other design choices included the color, radius, and surface properties of the point charges. We chose to adopt the common convention of red and blue positive and negative charges, respectively.

The radius chosen varied according to the scale of the scene, but was chosen to be large enough to grab and see clearly, but not so large as to unnecessarily obscure the space. A slight specularity gives the particles a natural physical look.

Dissipative Dynamics The dissipative dynamics of the system are defined by the dissipation term in Table 4.1. The formation of stable configurations, related to how molecules form, requires dissipative dynamics. Without dissipation, energy is conserved and particles can never come to rest. On the other hand, if dissipation is too great, then the dynamics are slow and the user must wait a long time to see the outcome of her actions. Therefore, the choice of a dissipation constant is more related to usability than faithfulness to reality, although the inclusion of a nonzero amount of dissipation is necessary to bring out important physical phenomena. Accordingly, we chose values that created a favorable tradeoff between these two. In a future version of the playground, it would seem useful to implement a method for speeding up or slowing down time, in order to experiment with a wider range of dissipation parameters in an expedient way.

Electric Field Lines One of the driving elements of the vision for electrostatic playground was to visualize electric field lines evolving in 3D. Associated with this is the challenges that field lines occupy 3D space densely, and they don't necessarily have a beginning or an end (i.e. they begins and/or end at infinity). A charged particle is, by definition, a source or a sync for electric field lines, and the chosen design leveraged this fact by propagating field lines forward or backward from the point charge (depending on its parity, positive or negative). We chose to originate such lines at 6 points on each charge, which corresponded to the x, y, and z axes (i.e. $(\pm r, 0, 0), (0, \pm r, 0), (0, 0, \pm r)$ where r is the chosen radius of charge sources) in the starting orientation. In the future it would be fruitful to experiment with different methods of defining which and what quantity of field lines to visualize.

Constant Electric Field As mentioned above, representing field lines in VR is challenging because they permeate space. In representing field lines that start and end at the charges, we leverage the fact that the most important and determining characteristics of the field lines in the rest of the space are those that are close to the particles. When we represent a constant field that permeates the space near and far, this solution does not apply. Naively applying an anal-

ogous approach would lead to filling the space with a large number of parallel lines. If the lines are dense, then it becomes difficult to see anything else; if they are too sparse it becomes less natural to interpret them as something that is actually everywhere. The solution implemented in this project uses moving translucent arrow that are spawned at random and constantly changing locations. This seems fairly effective because it gives the notion that there is something uniform happening throughout the space, and it doesn't seem to privilege any particular point or area. It also allows us to use the speed of the arrows to indicate the strength of the field. The down side, which can be misleading, is that the these moving arrows could be interpreted as physical entities; that is, something that something else could collide with. The learner needs to interpret them as symbolic indicators which are not physical objects themselves. It seems that this becomes clear after a brief amount of interaction in the space, so we found it to be adequate for the purposes of the project.

Implementation of Numerical Simulation

The numerical simulation necessary to power the Electrostatic Playground involves a standard explicit iterative method for discrete-time ordinary differential equations – the Runge-Kutta method [But16]. Thanks to Professor John Belcher, we were able to adapt existing software, which was written for making animated videos of electrostatic dynamics, in order to create an immersive, interactive 3D experience.

Equation	Description
$f = \frac{q_1 q_2}{r^2}$	pairwise forces
$f = -bv$	dissipation
$f = \frac{k}{r^p}$, $p = 5, 7, 9$	"Pauli force"

Table 4.1: Basic Properties of Simulated Electrostatic Dynamics

Lessons Learned

Some lessons learned concern (1) initial findings on the usefulness of visualizing and interacting with physical phenomena in VR, (2) the design of instructional environments, and (3) design considerations for learning-focused simulations.

Insightfulness Validated   Accompanying the desire to visualize and interact with physical phenomena in a new way was the question of whether it is truly valuable to do so, or merely fun and novel. Anecdotally, the answer is a resounding "yes." Having a dynamic 3D, sensorimotor experience, which must otherwise be simulated in the mind's eye, seems to yield a great deal of intuition. In the experiment presented in Section 5.1, I will attempt to prove this rigorously.

Low Ceiling vs. High Ceiling   The interactive experience, as it was informally piloted, was minimally structured– using a typical "sandbox" approach. Even as the prototype succeeded at validating the insightfulness of the experience, it failed somewhat in providing an arc of engagement that left participants feeling satisfied. A typical reaction after a few minutes was along the lines of "Did I try everything? Is there anything else I should try?" This highlighted an important aspect of the exploratory learning instructional approach, related to the design principle Resnick et al. called "low threshold, high ceiling, wide walls" [RMN+05]. This refers to three aspects of freedom and autonomy that need to be afforded to support creative thinking and learning; in particular, being able to get an easy start, having many choices about what to do next, and to support increasing sophistication. It was on the last of these that this project was somewhat unsuccessful. There was indeed an incredible amount of freedom as to how one could place particles in the space, but not enough opportunity to build up sophistication.

Fact, Fiction, and Feasibility   Another way of looking at the plethora of design decisions that were required to make the Electrostatic Playground usable is in terms of fact, fiction, and feasibility. The interaction we built to communicate *facts* about the electrostatic dynamics of dissipation and inverse square laws required implementing a set of *fictions*, like giant billiard ball-like particles that float in space. If we were to more accurately represent very small particles, we would need to render electron clouds, which would distract significantly from the principles, and also move away from the region of *feasibility*. The use of moving arrows is another example of a tricky design choice related to fact and fiction. The overall takeaway is that choices must always be made about where to be accurate and where to introduce fictions, due to considerations about what is accurate, salient, and feasible to build. It is extremely important to be mindful about this– it is very difficult for the environment to explain to the learner directly which aspects of the representation are fiction. If some of the fictions are interpreted as facts, this does a great disservice to the learner.

The project presented in this section is built around an interaction of direct 3D manipulation of parts of the relevant system (charges). Not all physical systems admit the direct manipulation of parts as a meaningful form of interaction. In the next prototype, I explore how interaction in 6 DoF can be used in another setting.

4.1.2 BrainVR: Exploring a Retinal Circuit with Paths of Light

This work was done collaboratively with the following people: Alex Norton and Amy Robinson (neural circuit content design); Max Rose (3D modeling); and Daniel Citron (visual design and interaction design). The 3D neuron models were reconstructed by users of the EyeWire[1] platform, and the behavior of the neural circuit was implemented as described by Kim, et al. [SKJGZ+14].

BrainVR was inspired by the idea of interacting with and visualizing a neural circuit in a physically meaningful way. It is a focused exploration targeted at a general audience, that combines exploration, capture, and review into one interaction. In particular, it uses of 6DoF spatial interaction as a way of interactively exploring the dynamics of a simulated system. The key mechanism involves looping the interaction, giving it the simultaneous purposes of exploration, capture, and review.

The learning goal of the experience is to convey the concept of a neural circuit that directly detects information about objects in the world. The particular circuit is from the human retina, which is responsible for detecting the direction of motion of moving objects [SKJGZ+14]. The older, more "naïve" view of the brain and the retina would have suggested that the retina pipes raw data to the visual cortex, and the visual cortex needs to figure out everything from whether there are objects at all to how those objects are moving. The discovery of this low-level neural circuit indicates that complex "preprocessing" happens directly in the retina.

Equipped Explorer Affordances

The learner is exposed to a three-minute audio-visual narrative showing the 3D structure of the brain and the eye, while an audio track describes relevant basic facts and figures. Visual images are shown in Figure 4-2. At the conclusion, one instance of the direction selectivity neural circuit

---

[1] https://eyewire.org/

79
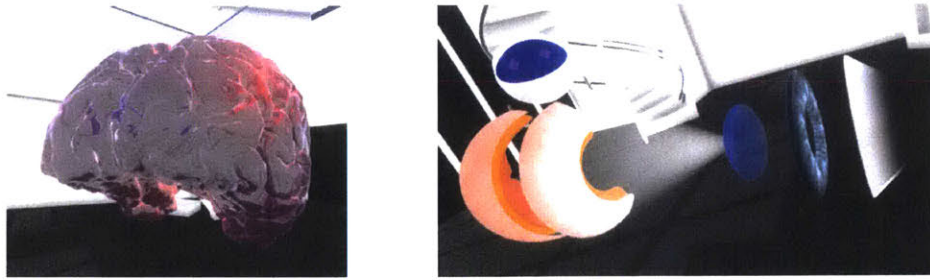
Figure 4-2: BrainVR Audio-Visual Narrative: 3D Brain and Exploded Eye View
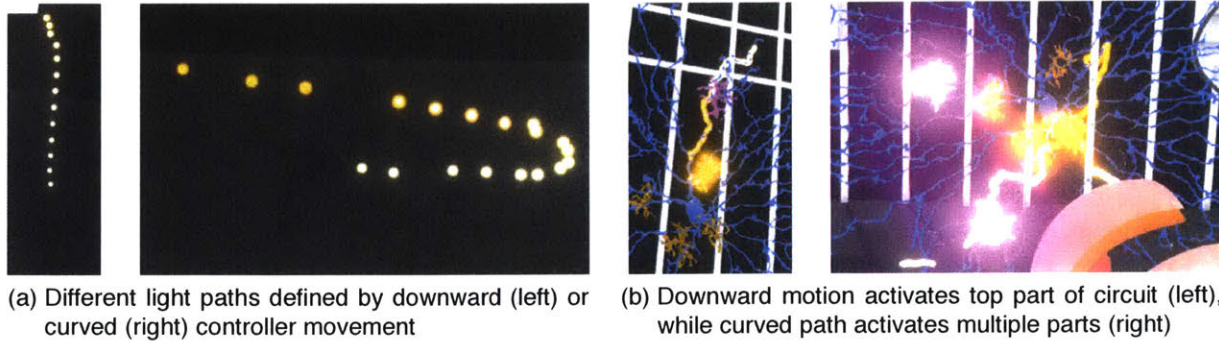
(a) Different light paths defined by downward (left) or curved (right) controller movement

(b) Downward motion activates top part of circuit (left), while curved path activates multiple parts (right)

Figure 4-3: User Interactions for BrainVR

is scaled to human size and placed roughly three meters away from the participant.

Exploration The participant holds a light source that can be used to record a repeating animation. For example, tracing the shape of a spiral while holding down the trigger would lead to a point of light repeating that motion over and over in a loop. The circuit is assumed to react to the path, detecting either left-right or up-down motion, and activation is represented by the entire neuron lighting up. This premise is introduced step-by-step, as the participant is instructed to try an up-down motion, then a left-right motion, observing how the circuit reacts in each case. Finally, the participant is free to experiment with any path of any length, and observe how its properties are detected in the circuit as its different components light up. These interactions are shown in Figure 4-3.

Capture and Review The looping interaction allows the participant to observe the same thing multiple times, utilizing the capture mechanism to enable review. Without looping, the participant would have to attempt to reproduce the same behavior in order to try to observe different properties of it. It would be more difficult to know whether differences have been observed, and if so whether

they must be attributed to variations in the participants' repetitions.

Observations and Lessons Learned

In this application, the exploration mechanism is explicitly tailored to support review as a primary method of making observations and deepening understanding. Repeatedly viewing the effect of the captured path frees the learner's working memory to focus entirely on understanding, while at any time allowing her to inquire further by recording a new path. This tool is generalizable and applicable to other scenarios.

A possible improvement might be to employ the contrasting cases method [GP92, SHA15] by allowing learners to contrast different recorded paths, either in sequence, or next to each other in space.

4.1.3   Neuron and Safari: Spatial Hyperlinks for Capture and Review

This work was done collaboratively with the following people: Daniel Citron (Neuron) and Hisham Bedri (Safari).

This section presents two applications that investigate the idea of spatial hyperlinking. In the first, Neuron, the learner holds a 3D model of a neuron, and is able to explore its structure by moving her head and hands, and making it larger or smaller. Vantage points can be saved and restored using an array of *Perspective Panels*. In the second, Safari, the learner flies through an environment populated with large 3D models relevant to biology (a human heart and a cell), and can take pictures and sketch notes. The pictures in this *spatial scrapbook* can be used as links to transport the learner back to the locations from which they were taken.

Motivation

Picking up on the line of thought on note-taking in VR presented in Section 2.2.5, I ask the question: what might active, synthesized and generative notes look like in VR? To begin, traditional handwritten note-taking is a possibility, assuming it can be captured and displayed at sufficient resolution, as shown in e.g. [PTW98]). Consider, however, that experiences can be recorded and played back in a straightforward way in VR. That is, a certain form of "verbatim notes" can be

made available without any attentional effort. This frees up more attentional resources to devote to comprehension and generative note-taking.

I propose considering visual snapshots as a conceptual basis for minimal verbatim notes. Different kinds of snapshots can be captured, and all of them can be used as hyperlinks. 2D snapshots are a familiar, lowest-common-denominator way of capturing visual information. Because the entire 3D environment already resides on the computer, 3D snapshots and notes are also a possibility. These could take the form of static snapshots of a scene to be revisited later, accompanied by one or many camera positions that the learner finds useful or enlightening. Traditional notes can be tied to entire scenes or to specific camera positions. When the 3D content is animated, the possibility of temporal hyperlinks arises. Snapshots may have multiple representations: 2D images small (thumbnails) and large can be embedded in 2D notes, or 3D snapshots could be used in a similar way. Animations can be represented as a series of keyframes that may be presented in parallel in space. One consideration is that it may be of significant value to design notes to be easily viewable on mobile devices and allow the review process to be more portable. 2D snapshots and hand-written notes do have this property, and this is a major reason to consider them as a central building block for a VR note-taking system.

As a minimal exploration of this idea, in this section, I propose an approach leveraging "hyper-linking" as a rapid form of active note-taking. This means using a system affordance to choose a location within the learning space as a point of reference to access later. Since this process is active– deciding and declaring that a moment is significant– we argue that a benefit associated with generative note-taking will be attained. Subsequently, when hyperlinks are reviewed, the content is accessible in its full original detail, allowing the learner to reap the benefits associated with verbatim note-taking. Synthesized notes can contain multiple hyperlinks, and the corresponding snapshots can be annotated with sketches and handwriting. Hyperlinks that are not embedded in notebooks may also be useful – similar to post-it notes used to mark important pages or chapters in a textbook.

Two prototypes presented here explore this idea of spatial hyperlinking by capturing perspectives on an object and locations in space, respectively.

Equipped Explorer Affordances

These two prototypes cover the purposes of exploration, capture, and review.

Exploration   The first of the two prototypes, Neuron, supports exploration of a physical object (a neuron) that is positioned and oriented based on the position of one of the handheld controllers. The object is a physically accurate neuron based upon a 3D reconstruction. In the second of the two, Safari, the learner is positioned on a moveable platform with a series of buttons that activate different tools. One of the tools is used for locomotion, allowing the platform to move continuously in a direction determined by the movement of the controller. Flying around to different locations allows her to gain novel spatial intuitions about these structures.
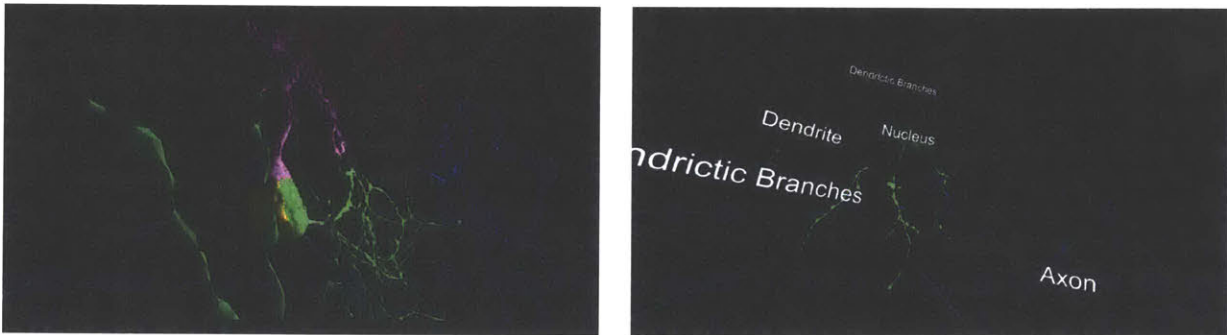
Figure 4-4: The BrainVR environment allows learners to explore 3D neurons. Labels for neuron parts can be displayed.

Capture   In the Neuron application, the learner can use the system to save an insightful perspective on the object. Recall that she is exploring a 3D model of a neuron. The insightful perspectives might include simply the names of the different parts of the neuron, as shown in Figure 4-4, or more subtle details such as how certain dendrites branch in comparison to others. She wishes to save and share these insightful perspectives – which are defined by an orientation, camera position, and level of zoom for the object. A button on the second controller allows her to take a snapshot of the perspective. The perspective is then added as a graphical thumbnail to an array of Perspective Panels, as shown in Figure 4-5. In the *Safari* application, shown in Figure 4-6, the learner can use the system to save locations and perspectives within the complex landscape she is exploring. She can take photos, sketch on the photos, and arrange them on a canvas or book that travels in front of her like a portable drafting table.

83
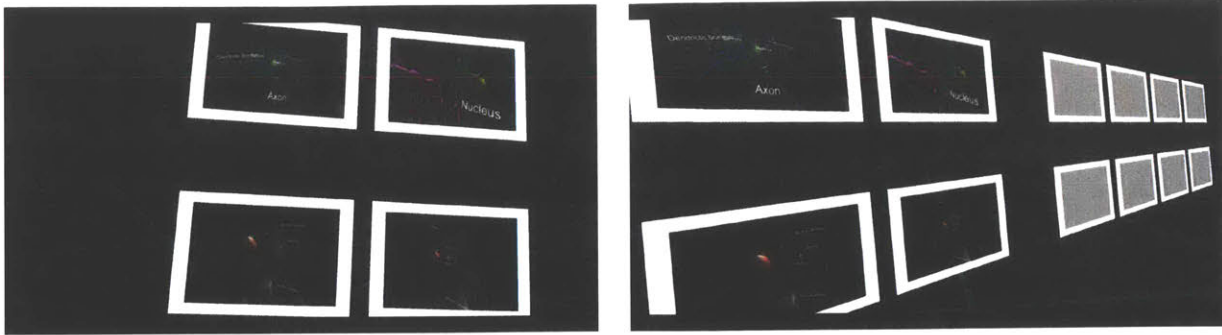
Figure 4-5: In Neuron, the Perspective Panels capture object perspectives. They function as hyperlinks to restore a perspective by reorienting the object (a neuron, in this case). New perspectives populate the gray squares.
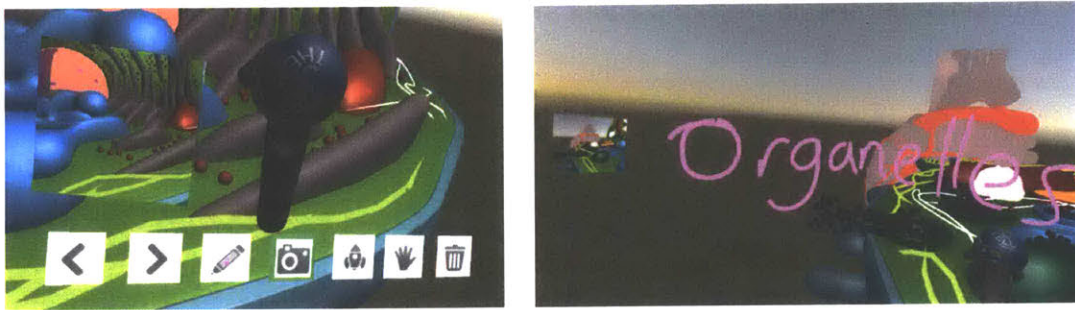
Figure 4-6: In Safari, the spatial scrapbook captures camera perspectives from different locations in a space or model. Thumbnail images function as spatial hyperlinks.

Review   Recall that the learner using the Neuron application collects snapshots as an array of panels. This array facilitates review when the user touches one of the populated panels. The object rotates visually so that the current perspective matches the perspective from the time of capture. This way, the Perspective Panel acts as a memory cue to recall where the insight was had, but restoring the 3D world to that state provides a much richer cue that also supports further exploration. In the Safari application, viewing the spatial scrapbook with annotations and thumbnails is one form of review. In addition, the photo thumbnails collected in the scrapbook can then be used as spatial hyperlinks to return to the location in the 3D environment from which they were taken.

Implementation

The Neuron application was built for the Oculus DK2 and Sixense STEM system, while the Safari application was built for the HTC Vive.

Lessons Learned

Our initial trials showed Neuron's Perspective Panels to be a very effective and intuitive way of saving object perspectives for later review. It represents a reusable tool that can be applied in many other settings. Piloting the spatial scrapbook in Safari showed this provides an effective way of organizing information gathered through exploring a large, complex landscape with details in different locations and at different scales.

One shortcoming of the spatial scrapbook linking mechanism is that abrupt transitions between locations are jolting and make it difficult for users to infer spatial relationships between locations. Future work could explore solutions involving eased linear motions, and visual guides to indicate the path between locations. Providing a better method for learners to understand the spatial relationships between different locations they have visited is an important challenge to address.

Moving forward, it will be important to evaluate different note-taking methods for relative effectiveness, and also effectiveness compared with 2D notes. Issues to consider are: efficiency, since flipping quickly through 3D locations might be jarring, and/or take time to adapt to; the use of notes in non-VR settings, as raised in the Section 3.3; methods for capturing text and symbols, since it is not clear when different methods like hand-writing, speech-to-text, or virtual keyboards might be appropriate; and methods for integrating 2D learning materials into the 3D environment. The last of these will impact methods of note-taking. It is an important case to consider, since there is already such an abundance of materials in 2D that are both high quality and viewable on any platform. Clever ways of displaying these in relation to interactive 3D content are sure to be highly valuable.

Summarizing what was accomplished by these two note-taking prototypes, I provided two generalizable interface tools that can facilitate capture and review in a variety of settings: perspective-based hyperlinks and location-based hyperlinks. These were effective for the purpose they were intended, and raised a number of interesting issues and areas for future work. These involve methods for relating notebook materials to the environment around them, other relevant learning materials, and each other. All of these will be important for making the most effective possible systems for capture and review, that leverage the ability of the virtual environment to provide effortless

verbatim notes that augment active, synthesized notes.

4.1.4 TerrImmerse: Exploring Terrestrial Data Using Multiple Representations

This work was done collaboratively with the following people: Jonathan Stets (volumetric surface view), Max Rose (visual design and interaction design), and Wiley Corning (interaction design).

The TerrImmerse application allows users to explore and annotate complex 3D datasets. In this case, the data was terrestrial density data used to understand the topology of geological structures. Challenges exist in segmenting and verifying this segmentation of this kind of data. This process requires working with two representations of the data– raw data, and segmented data. Segmenting and verifying the segmentation of data both require an understanding of the topic, most importantly knowing how to identify problems. This interface was envisioned as a tool for training and practice of work with this kind of data. Because volumetric data and working with multiple spatial representations are broadly applicable, this is seen as on example that is representative of many.

Equipped Explorer Affordances

This application covers the categories of exploration, capture, and review.

Exploration   As mentioned above, the TerrImmerse application deals with two spatial representations. These are the volumetric view and the surface view. The following affordances allow datasets in either view to be explored: (1) head movement, (2) scaling, (3) translation, (4) rotation. Figure 4-7 illustrates the functioning of each of these tools.

The two representations related to each other as follows: the dataset represents geological density in a volume of earth, and the assumption is that, within the volumetric data, there are curvilinear planes of constant density that correspond to geological layers that have built up over the millennia. This application is allows exploration of these two representations in relation to one another, assuming that surface generation has already been done outside the application. The purpose is for learners to understand the relationship between the layers, and for experts to be able to verify the accuracy of the layer segmentations given.

(a) The translation tool allows the user to move the dataset to a new position

(b) The rotation tool rotates the dataset about the z-axis

(c) The scaling tool is used to make the dataset larger or smaller

Figure 4-7: Core Exploration Features in TerrImmerse

(a) Surface view planes may be shown or hidden individually

(b) All may be shown or hidden at the same time
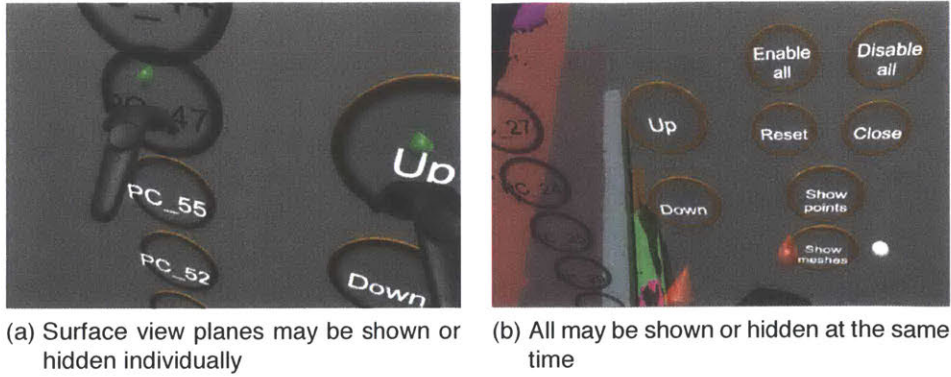
Figure 4-8: Surface View Exploration Features in TerrImmerse

(a) Moving front cross-section in the rear direction

(b) Moving three orthogonal cross sections simultaneously
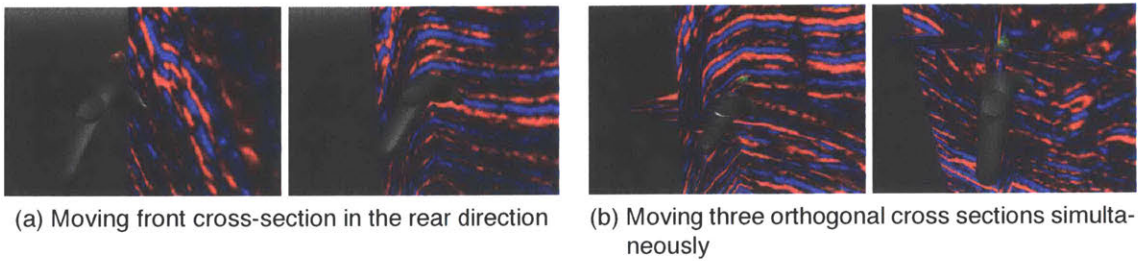
Figure 4-9: Volumetric View Exploration Features in TerrImmerse

The surface view allows subsets of surfaces to be shown or hidden. This can be accomplished either by selecting a surface directly in space, or using a list of checkboxes. An efficiency feature in the interface design allows the user to drag across many checkboxes to toggle them with a single trigger-down. The exploration tools using surface view are illustrated in Figure 4-8.

The volumetric view allows cross-sections of data to be explored along three axes. The *xy, yz,* and *xz* planes may all be independently dragged along their orthogonal coordinate axis to reveal different cross sections of data. Sets of two or all three of these planes may be moved simultaneously by positioning the controller at the geometric intersection of the desired set. This functionality is illustrated in 4-9.

Finally, juxtaposing the surface view with the volumetric view gives further insight into how the representations relate to each other, and where there might be ambiguities or errors. This is enabled by a quick, always-on gesture – using the thumb to swipe left and right on the controller touchpad.

(a) Measurement tool and endpoint marker

(b) Measurements are visible in either view

(c) Trigger-and-move action carries out measurement

(d) Multiple segments form more complex annotations, e.g. closed shapes and distances from one point to many points
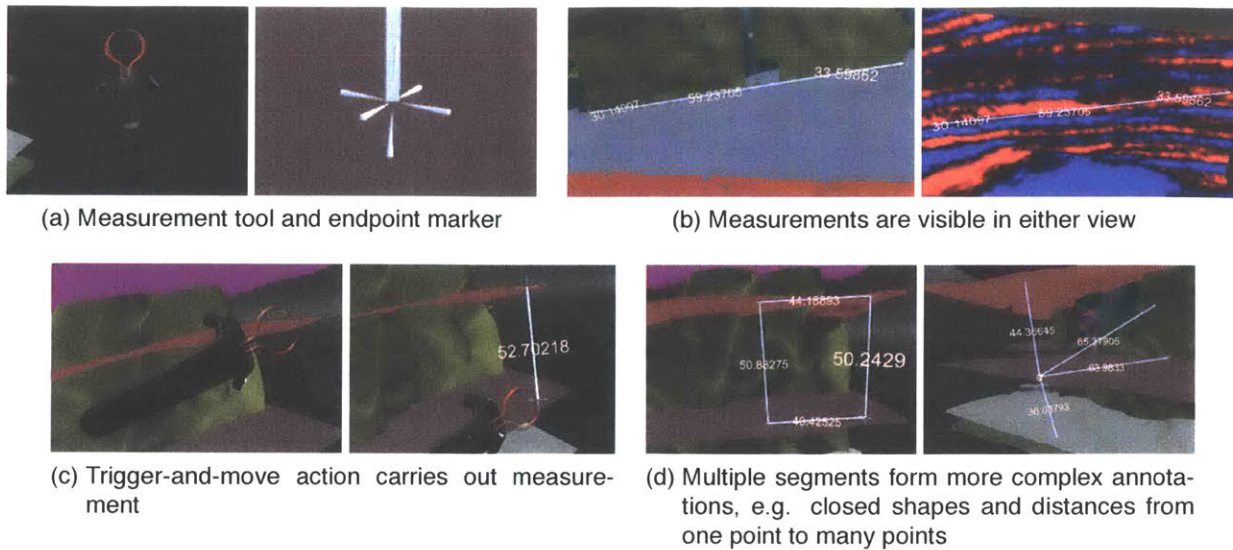
Figure 4-10: Measurement and Annotation tool in TerrImmerse

Capture   Accurately identifying topological properties of geological structures is only an intermediate step between collecting data and drawing general conclusions. The output of this step needs to be annotation, which can be either in the form of formally modifying numerical data (e.g. in geometric representations of the said curvilinear planes), or in observations make as sketches or descriptions, to be used for human analysts at later stages. This application allows for annotations to be made directly in 3D space with the data using the VR system's handheld controllers.

The measurement tool functions consistently in both views. It allows for piecewise linear curves to be drawn in 3D space. Each segment is labeled with its numerical length. This allows for lengths to be measured, and areas or volumes estimated. Such lines can be used to roughly highlight an area of interest, or to precisely mark points or segments. Using the measurement scaling, arbitrary levels of precision are achievable with respect to the original dataset. In the example, it is easy to exceed the resolution and precision of the original dataset through the use of scaling. The functioning of this tool is illustrated in Figure 4-10.

Review   Review is done by inspecting annotations created by the learner or someone else. These can be easily saved and loaded along with the data. Switching back and forth between the two representations with annotations remaining constant allows the learner to clearly see specific insights that are being highlighted.

Lessons Learned

There were multiple lessons learned from this prototype, primarily concerning the use of multiple representations, affordances for browsing complex spatial data, and the the need for Reality Integration.

Firstly, juxtaposing multiple spatial representations was very successful. Anecdotally, seeing these representations stereoscopically overlaid and exchanged seemed to enable clear insights into complex phenomena.

Second, there were some specific challenges that we encountered relating to efficiently arriving at the desired view. One is that defining an appropriate scaling curve is nontrivial. Using linear scaling, the dataset either scales much too fast when it is small, or much too slowly when it is large. Polynomial scaling helped but need to be carefully calibrated. It is possible that this should be calibrated for individual users' body size, so that moving the arm from outstretched to retracted has the same effect for everyone (e.g. smoothly doubling the apparent size of the dataset). Next, selecting surfaces to be hidden or shown from a long checklist was inefficient, and a method like direct selection would be preferable. This would require a form of ghost or wireframe representation to allow hidden surfaces to be shown again.

Finally, in this particular use case it is clear that data must come into and back out of the system as part of a practical usage, and so *Reality Integration* (as introduced in Section 3.3 comes into play. New data must always be brought in, either to provide diversity to learners, or to export the insights captured by experts. Our prototype used a rudimentary, manual process for importing data that would need to be made automatic to make real use of such an application.

4.1.5   Section Summary: Exploration, Capture, and Review

In this section, I have presented prototypes that are most relevant to the single-user purposes of exploration, capture, and review in the Equipped Explorer framework. These proof-of-concept implementations have, by and large, confirmed the validity of the basic design ideas that inspired them, and produced generalizable tools that can be applied across many contexts. These include the consideration of: analysis of fact, fiction, and feasibility when designing simulation-based learning

environments; using free-hand input to interact with simulated systems that cannot be manipulated directly; the use of hyperlinks to capture and review perspectives and locations during exploration; and a set of interaction tools for exploring and annotating spatial datasets with multiple representations. In the next section I will discuss prototypes that add collaboration and communication to the list of purposes.

4.2　Communication and Collaboration

In the previous section, the focus was on prototypes for single-user Equipped Explorer purposes: exploration, capture, and review. In this section, I bring the lessons learned and questions raised into context with the fourth purpose defined in Section 3.1, namely communication and collaboration (I will use these interchangeably; either use refers to this same purpose). Is it feasible for multiple users to collaborate in a shared virtual and physical space? Are there any unexpected challenges or opportunities? I will continue with the approach of developing general-purpose tools within the context of concrete prototypes. The three prototypes presented in this section explore the following: collaborative interaction with a simulation; collaborative 3D drawing; and communication and collaboration in a creative open world landscape.

4.2.1　Body Quest

This project was realized with the help of the following people: Wilhelm Weihofen (biochemistry content design), Max Rose (animation), Wiley Corning (interaction design), and Theji Jayaratne (video production).

The Body Quest project explores the use of interactivity in the context of biology and chemistry. One tension in the production of interactive content related to detailed process knowledge is that it can be difficult to make it interactive. That is, if a linear process depends on a sequence of things happening in a particular way, then user interaction will most likely just prevent the key phenomenon from occurring. While this might have some value – illustrating the sensitivity of the environment, for example – if the main idea is to understand a sequence of mechanisms, then preventing those mechanisms from being presented at all is a problem.

The design approach that I propose has two components: (1) isolate each mechanism and create from it an interactive scenario, using a visual narrative to frame the scenario within its context, and (2) illustrate structural components of the system using in-situ annotation, and interactive exploration that is outside the scope of the animated dynamics. The user interaction will allow the user to control the environment and the dynamics in certain ways that are "non-physical," but this can be designed in a way that makes the distinction between the natural and interactive elements

92

easily understandable. In this example, I will describe how I applied this approach to create two interactive experiences within one scenario.

The core mechanism that this project illustrates is how the flu virus employs a protein to destroy the mucus chain structures that the human cell uses to defend itself. These chains are long, repetitive structures, and the viral protein repetitively applies the same attack to break them down, one unit at a time. In the real system, a variety of fluid dynamic processes affect how the virus naturally approaches the cell. In order to establish what the "natural" interaction looks like, I apply a linear narrative/non-interactive framing animation. The animation is immersive, so it supports inspection from any perspective or location in the room-scale space, and may be paused, but does not permit manipulation of the individual elements. Some images from the animation are shown in Figure 4-11.
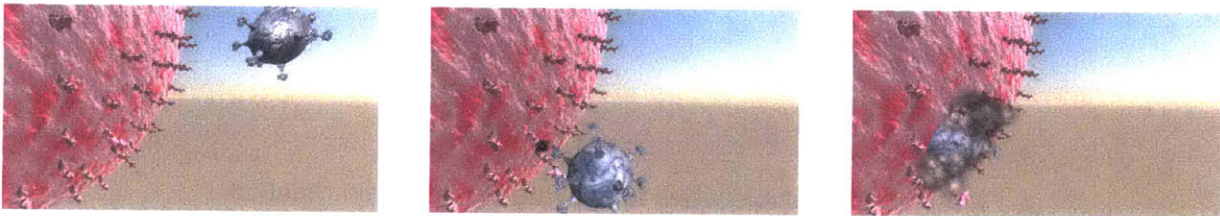
Figure 4-11: Virus approaches, attacks, and enters the cell

Equipped Explorer Affordances

This project addresses two Equipped Explorer purposes: exploration and communication.

Exploration   The goal of this scenario is for the learner to gain insight into how the flu virus breaks into human cells, and how the implementation of this mechanism is achieved using protein folding. First, a non-interactive animation of the virus breaking into the cell provides a context for subsequent interactive segments. Next, the virus and the cell are presented next to one another, and the two different interactions involve (i) activating the key chemical reactions interactively, and (ii) inspecting the structure of the protein itself using multiple representations.

In the interactive presentation of the system, the user can either move a mucus chain towards

the protein, or the protein can be moved towards a mucus chain. Because the key interaction is spatially local (i.e. it only matters that the protein and the mucus chain are aligned with each other), we can essentially freeze the system's time-varying dynamics that would cause this alignment to happen on its own, and allow the user to manually move individual elements into contact. This dual form of interaction is shown in Figure 4-12. A hypothesis here is that this distinction is easy to understand when shown in combination with the animated scenario framing that shows how this contact happens naturally when all the involved components are floating around in the cell's cytosol. That is, the user can cleanly substitute the natural fluid-induced movement shown in the animation with her manual manipulations in the interactive mode to explore the physical system's properties. Thus she gains the ability to develop understanding through exploration and experimentation without being confused about the causal aspects of the real system.
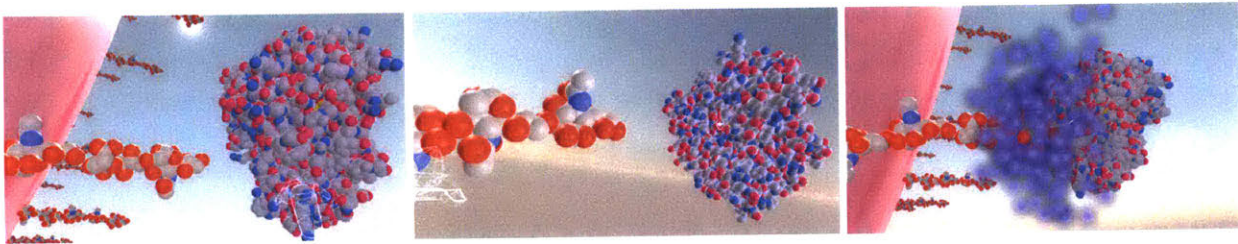
Figure 4-12: Multiple ways to activate mucus chain reaction

In order to give more context to the physical form of the viral protein – and emphasize that this is the product of a process of DNA transcription and protein folding – we introduce multiple views of the viral protein that emphasize different characteristics. The three views, shown in Figure 4-13, are described below:
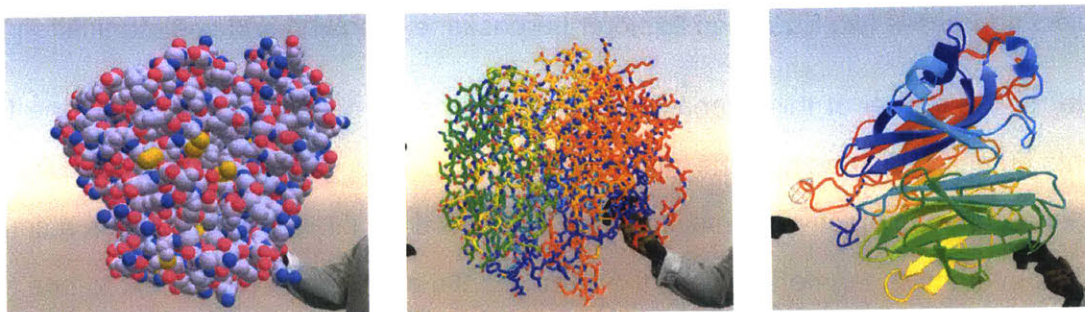
Figure 4-13: Views of Viral Protein

(i) **folding path ribbon**, which uses a 3D ribbon representation. A flat ribbon that morphs into

94
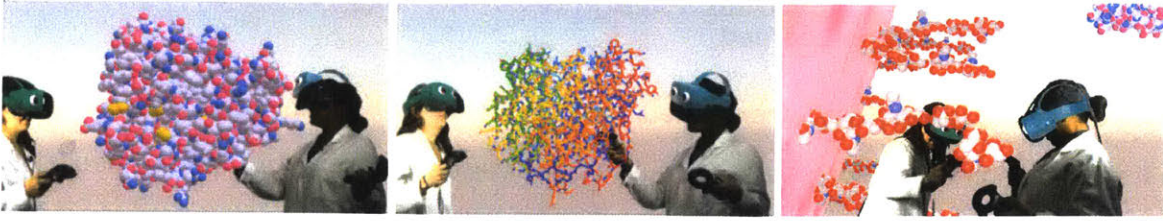
Figure 4-14: Multiple Users in Body Quest

a flat arrow at intervals, to show the direction of transcription and folding. A continuously varying rainbow coloring allows the viewer to estimate the location of a given point on the ribbon with respect to the start and end of the folding path.

(ii) **small ball-and-stick representation**. This representation makes all atoms and amino acids individually visible, while keeping the entire structure transparent. The adds clarity to the concept of amino acid transcription and folding, especially when juxtaposed with the folding path ribbon.

(iii) **large ball-and-stick representation**. This representation emphasizes the nature of the outer surface of the folded protein, while still keeping the connection to individual atoms immediately apparent. The outer surface (including the active site cavity) is what causes the protein to perform its catalyzing function – the geometry of the active site fits in a "lock-and-key" fashion with the end of the cell's mucus chains.

Communication and Collaboration   The Body Quest application was designed with communication and collaboration as a core part of the experience. There is no built-in presentation of facts related to the content, making it difficult to understand without the guidance of an expert. The following four key collaboration affordances were implemented to support the usage scenario. These are illustrated in Figure 4-14:

- Gaze awareness and shared space.

- Simultaneous interaction with simulation.

- Multiple representations and shared control of objects.

- Question answering through pointing.

Observations and Lessons Learned

In Body Quest, I successfully demonstrated a method for learners to interactively explore a process that would normally be presented linearly for passive consumption. An expert guided a learner through the content, with no difficulty on the part of the interface design: easily pointing to establish reference points for discussion, and interacting simultaneously with the simulation-driven environment. This shows that our straightforward method of putting people in a shared virtual and physical space, represented with minimal avatars, provides a solid starting point for developing a variety of collaborative applications. Although the scenario itself was effective, Reality Integration would appear to require (1) a larger database or more versatile simulation environment to handle numerous such examples, and (2) the integration of factual content including narrative explanations and diagrams to support the learner in going from the concrete to the abstract.

4.2.2   CocoPaint: Collocated Collaborative Painting

This work was done collaboratively with Hisham Bedri and Ronen Zilberman, who contributed to the implementation and interaction design.

The goal of CocoPaint was to preliminarily explore the feasibility and user experience of same-time, same-place collaborative freehand drawing in room-scale VR. None of the projects presented thus far have delved deeply into the *creation* aspect of exploration, which is so central to constructionism, and important to constructivism in general. Drawing in 3D is a constructive and creative activity, and it can be used across many learning contexts. Therefore this prototype at the intersection of creation and collaboration covers an important area, as I continue to survey the activities of the Equipped Explorer.

Equipped Explorer Affordances

The purposes supported by this application are exploration, collaboration, capture, and review.

Exploration, Capture, and Review   Drawing and painting are powerful media for expression in 2D because they are simultaneously forms of exploration, capture, and review. What has just been done by the user is captured on the page, and one alternates between reviewing and further

96

(a) User avatar   (b) Users collaborating

(c) Selecting a color   (d) Collocated users
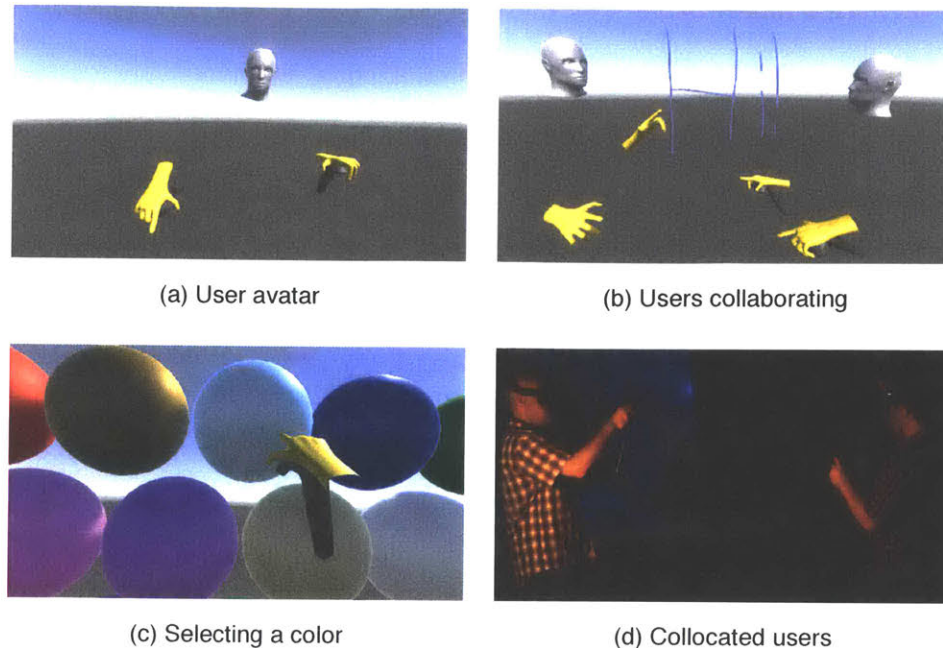
Figure 4-15: Initial CocoPaint Design

creating through the authorship process.

In the initial iteration, shown in Figure 4-15, users could select a color, and paint with a single line thickness in 3D space. Color selection was done by touching the index finger of a hand avatar to spheres of different colors, located statically in the virtual room space.

Collaboration The head and hand avatars can be used for non-verbal communication, such as pointing and nodding. This can be used for coordinating and planning next steps. The scale of the paintings allows people to work on different parts at the same time. The space is large enough so that participants can choose whether or not to work together.

Observations and Lessons Learned

Considering the exploration affordances, the flat line geometry used for painting was very limiting: its appearance from the side was counter-intuitive and unpredictable, and the fact that the color varied significantly depending on viewing angle was confusing.

Concerning collaboration, a number of important lessons were learned about the design of minimal avatars:

- Simple avatars are highly expressive (see Section 5.2 for a detailed treatment of this topic).

- In order to facilitate simple hand gesture communication, it is necessary to select a versatile hand posture. Examples that did not work: pointing index finger (initial prototype), hand gripping controller.

- Adding eyes to the headset avatar had a profound impact on social presence (differed between initial and revised prototypes).

- Aligning the position of the head avatar with the real head is critical for head gestures (nodding, shaking, etc) to be interpretable (differed between initial and revised prototype).

The experience of being expressive through gestures and painting in a shared virtual and physical space was highly compelling for the users who tried it. The conclusion was that this is a powerful basis for applications where creation is a core component. Because of the success of this prototype, we began a series of further iterations that are recounted in the next section.

4.2.3 Iterative Contextual Design for Collaboration and Cocreation

In this section, I will first present two stages of iterative design of an application for collaboration and cocreation, extending and improving the interaction piloted in CocoPaint. Then I will discuss findings from piloting the application with many dozens of users. The first stage is a straightforward extension of CocoPaint where the interaction is situated in a realistic 3D environment, and the objects in the environment can be incorporated into users' paintings. The second stage extends the basic painting functionality with a diverse, open-ended set of creation and exploration tools and affordances, under the name CocoVerse. Finally I discuss findings from piloting the application with many dozens of users, in terms of prototypical behaviors and activities that they tended to engage in.

CocoPaint in Rome

On the strength of the CocoPaint prototype, I started a project with an industry collaborator in order to test the feasibility of this style of collaboration in an applied context. Their use case was facilitated collaborative design – a facilitator was responsible for setting up task and guiding the design
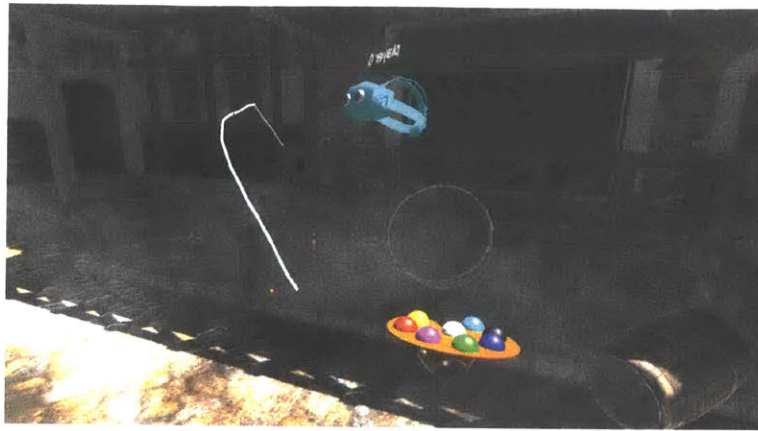
Figure 4-16: CocoPaint in Rome

activity for the participants, and it was critical to be able to combine the ability to collaboratively design with the ability to be immersed in an environmental context. This section describes the first iteration.

Equipped Explorer Affordances   The initial aim was to establish the effectiveness of immersion in an environmental context. This would be evaluated by the extent to which participants demonstrated and attested to inspiration related to the surrounding. To begin, we imported the CocoPaint collaborative painting user interface into a rich 3D environment – a recreation of a section of medieval Rome, as shown in Figure 4-16. This implemented the same Equipped Explorer purposes as CocoPaint above: exploration, collaboration, capture, and review.

Observations and Lessons Learned   The expert facilitator concluded that the combination of a rich virtual environment with sketching, including 3D objects that could be incorporated into sketches, would be highly beneficial for the use case. In order to further advance the project, we identified the following requirements:

- A workflow that would allow end users (in this case, facilitators) to input their own environments without programming or being assisted by developers.

- A method for capturing the designs generated in VR to be viewed and further discussed outside VR.

99

At this stage, we were also asked to provide guidance on the differences between 2D, 360, 3D/360, and 3D graphics content. After some dialog, we concluded that 360 images were the best suited to the use case. This was because of (i) the need for diversity of content and frequent new content and (ii) the technology and staff available to produce this content.

We implemented our next iteration in a new, rewritten application called *CocoVerse*, which will be discussed in the next section.

Initial Design of CocoVerse

This work was done collaboratively with Wiley Corning, who did all of the implementation, and contributed significantly to design decisions in the process.

The aim of the next iteration of CocoPaint, which different enough to warrant the new name CocoVerse [GCM17], was to introduce a richer set of creation affordances (fitting into the Equipped Explorer category of exploration).

Equipped Explorer Affordances The various tools presented here cover all of the different Equipped Explorer affordances. The initial design of CocoVerse meets the design requirements mentioned above, and addresses the limitation concerning the line geometry that was mentioned above in the discussion of CocoPaint. The basic user interface metaphor of CocoVerse is that of *tools* and the *tool belt*. Each hand holds one tool at a time, and that tool can be exchanged by activating the belt and selecting a new tool. The belt is activated by clicking in a region near the waist (defined with respect to the headset). These are shown in Figure 4-17.

The following tools covered our design requirements:

- **Hands**. These are used to point and gesture. Several hand postures can be activated, including pointing index finger and thumbs-up postures. This is a communication tool.

- **Brush**. The brush allows creators to draw in 3D. The configurable attributes of the brush include color, texture, and cross-section. This is an exploration tool that leverages capture and review as described in the CocoPaint section, and it uses the high-bandwidth spatial input modality.
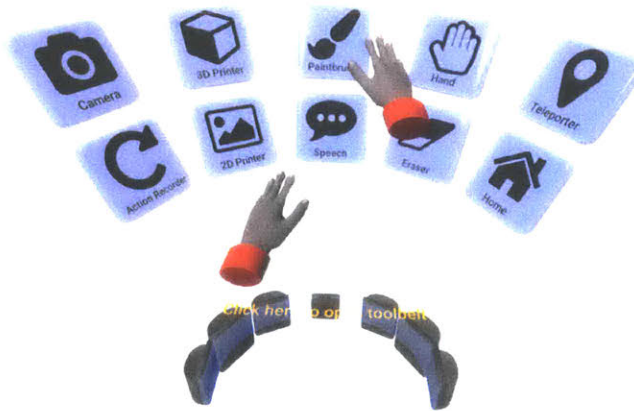
Figure 4-17: CocoVerse: Basic Design

- **Printer**. 2D and 3D variants of the printer allow creators to select an image or 3D model to place in the environment, with an adjustable size. A special 3D object allows users to place panoramic 360 images in the environment. This is an exploration tool similar to painting that is expressive in a different way. It uses the low bandwidth spatial input modality.

- **Camera**. This tool allows the creator to save 2D snapshots from the environment for later use outside VR. It is a tool for capture, and uses the low-bandwidth spatial input modality. The selection panel acts a tool for review, and enables further expression when previously captured images are instantiated in the environment.

- **Teleport**. Locomotion allows creators to utilize a large virtual space for ideation. This is an exploration tool that uses the low-bandwidth spatial input modality.

The workflow requirements mentioned above are matters of contextual Reality Integration, and we supported them in the follow way (illustrated in Figure 4-19):

- The facilitator can drop 2D and 360 images into folders in the application directory to import them into VR.

- Snapshots from the camera tool are saved in a folder in the application directory (outside VR).

101

Subsequent development for Reality Integration is described in Section 4.3. I reflect on the findings and lessons learned after the next section, which presents prototypical activities and behaviors.

Figure 4-18: CocoVerse Tools: (from top left) Brush, 3D Printer, 2D Printer, Teleporter, Camera, Hands

Fun Findings: Prototypical Activities and Behaviors in CocoVerse

The CocoVerse prototype application took my research team by surprise when we found participants using it for much longer periods of time— essentially not wanting to leave. Over the course of the application's development, we have had well over 100 users, and have been able to observe a number of prototypical activities and behaviors. In this section I categorize and catalog examples of these, and offer some observations about the contributing factors, and how they may generalize to other applications.

I divide the overall factors that appear to contribute the most to the engaging and fun qualities of CocoVerse into three categories. I will give examples below, after elaborating on them briefly here.

(i) *Novelty.* Some basic affordances produce visually surprising or rich experiences. Novel interactions are common discoveries that are engaging, even for solo users, and are primarily

(a) Images can be imported from outside VR: 2D (left) and 360 (middle, right)

(b) Snapshots can be placed in the world, or accessed outside VR

Figure 4-19: CocoVerse: Basic Workflow

enabled by one tool.

(ii) *Generativity.* The open and generative nature of the environment and tools seems to give rise to engagement. The modular aspect of tools that allow their abilities to be combined in interesting and unexpected ways.

(iii) *Collaboration support.* Users are excited by the ability to immediately share discoveries or creations, and directly engage with and influence others' experiences and avatars (e.g. by throwing objects at them). Collaborative interactions seem to fit into the three categories of *sharing*, *competition*, and *coordination*.

Now I will give examples from each of these categories. Note that these were emergent, rather than designed, phenomena. Figure 4-20 shows examples of generative and novel interactions, while Figure 4-21 shows examples of collaborative interactions.

Novel Interaction 1: Drawing knots. This seems to be a basic instinct for any user drawing in 3D for the first time. They draw curves and immediately begin to spiral around the other curves and draw "knots" that intertwine without self-intersection. It seems to be the most direct way to observe that the produced sketches are actually 3D.

Novel Interaction 2: Producing large things.   Both the brush tool and the 3D printer allow the user to easily create and manipulate objects that would be large and unwieldy in the physical world, but are of course weightless and effortless to deal with in VR. The draw of this experience would seem to be its difference from the physical world.

Novel Interaction 3: Erasing many things.   Once users have cluttered their spaces with many objects from the 3D printer, they are frequently incline to "clean up" the space using the eraser. Because the eraser can be set to a projective mode, many objects can be erased by standing in one place and scanning in different directions with the trigger held down. There is a small amount of haptic feedback when an object is erased as well. The combination visual decluttering and haptic feedback seems to be very satisfying for users, to the degree that they verbally comment on the novel and surprising satisfaction that it produces. It is possible that this could be attributed to the small amount of energy required compared with the physical world. Seeing a room full of clutter, or a tennis court with balls all around, a person has a certain physical expectation for the time and energy required to clean the clutter. VR provides the pleasant experience of decluttering using a fraction of this time and energy.

Novel Interaction 4: Peering through stained glass.   The brush tool can be configured to paint with a stained glass texture. This stock texture is implement in such a way that peering through it produces irregular and realistic refractive effects. Peering through large, curved and knotted 3D structures composed of virtual stained glass seems to be a novel and tantalizing experience, judging by user behavior and feedback.

Novel Interaction 5: Teleporting and standing in mid-air.   Teleporting is one of the most novel and empowering experiences in the virtual world, that diverges significantly from the physical world. Some people experience anxiety when standing in mid air, while others do not. Anecdotally, it appears that the vast majority of users have this experience only for a minute or two, and then completely forget about the novelty, or how scary the comparable experience would be in the physical world.

Generative Interaction 1: Erasing objects that contain other objects.   This is a common example of the capabilities of the CocoVerse environment that users are given to introduce them, and

104

encourage them to think in terms of trying out combinations of tools. Any concave, upright shape that is produced by the paintbrush can contain dropped or placed objects created by the 3D printer. A basket made using brush strokes can contain, for example, an assortment of fruit. The eraser tool can then be used to make the basket disappear, and as the rules of gravity dictate, the formerly contained objects fall to the ground. This seems to produce amusement and engagement.

Generative Interaction 2: Teleporting while painting. It is delightful to discover that paint strokes can be continued while teleporting. Using this method, users can create virtual paint strokes that span large virtual distances.

Generative Interaction 3: Rolling objects on the inside of hollow brush strokes. The volumetric brushes in CocoVerse create hollow interior spaces, and these can contain objects produced by the 3D printer. This is particularly entertaining when combined with the previous example: a pear or bouncy ball can held by or rolled down a long painted tube.

Collaborative Interaction 1: Sharing a creation or ability. Users frequently engage in solo explorations of the creation tools. They are often entertained by the visual or geometric qualities of the things they make: stained glass is fun to look through, giant bananas are humorous, and bouncing balls off of things is fun to watch. When they have done something fun, users will entreat other users to teleport to where they are, in order to have a look.

Collaborative Interaction 2a: Dropping objects on each other. In this example, one user drops things on or in front of another user. The idea is that they will be surprised or annoyed, and look straight up to find the sneaky second user using the 3D printer to drop things on them. The natural reaction for many users is to try to reciprocate and drop things on the other user. This qualifies as a *competitive* interaction.

Collaborative Interaction 2b: Painting on each other. In this example, one user paints on or around the other user to confuse and incapacitate him or her. She can't see anything until she either teleports away or erases the blocking paint stroke. This can often lead to a confrontation where users will use an eraser and a brush to walk around, erase defensively, and paint offensively. This qualifies as a *competitive* interaction.

Collaborative Interaction 2c: Firing a fruit cannon at each other. In this example, one users discovers that the fruit cannon can fire things at another user's face. He or she creates one or more fruit cannons and repeatedly fires them at the other user. This qualifies as a *competitive interaction*.

Collaboration Interaction 3: Playing catch. In this example, two users attempt to coordinate throwing objects to each other. The interface is not well-adapted to the precise timing required, so it is very challenging but not impossible. This quality seems to contribute to the engagingness of the activity. This qualifies as *coordination*. In another variant of this activity, one user stands downhill from another user, and attempts to catch rolling objects. This is easier than catching in mid air, since the objects need to be blocked. The paint brush can be used to create barriers that also block the objects, which users sometimes discover and enjoy.

Now I present observations about what aspects of the underlying implementation contribute to the ability of an application to support these attributes, and refer to these as *enablers*. These are aspects of the platform and user interface implementation that appear to have been critical to the emergence of the prototypical interactions named above.

Enabler 1: Modular tools and implementation. The modular tool implementation allows examples like those above to emerge organically. Viewed combinatorially, this gives rise to more possibilities than the designer could possibly enumerate or imagine. Applied to other environments or designs, this property is likely to produce similar results.

Enabler 2: Powerful physics engine. To give credit where credit is due, this application was built on incredibly powerful hardware and software provided by commercial vendors. What for previous generations of researchers and developers might have taken weeks or months can now be done in a matter of hours or even minutes: an odd-shaped object that rolls, or a glass texture that produces realistic refractive effects, for example. Much of the complexity that arises from the above interactions is indeed a consequence of this hardware and software. People enjoy the complexity of balls bouncing chaotically through uneven landscapes, and this requires almost no development time at all using modern graphics cards and development environments.

(a) Producing large objects (novel)

(b) Erasing many objects (novel)

(c) Making objects to contain other objects (generative)

(d) Erasing objects that contain other objects (generative)
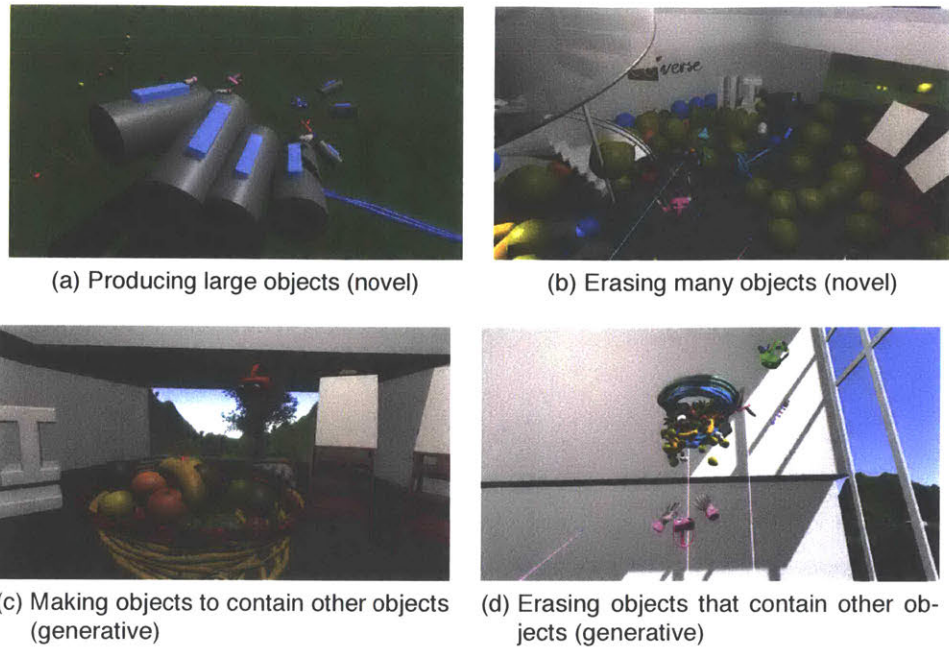
Figure 4-20: CocoVerse: Novel and Generative Interaction Examples

Discussion CocoVerse fits the overarching paradigm of the *sandbox* or *open world*, where the user is given the sense of being entirely in control of their experience– they are creating something themselves, as opposed to consuming something that was prepared for them. Reflecting on the above examples, it seems that one can attribute fun in a VR sandbox environment to (1) novel experiences that would not be possible in the physical world, (2) experiences that are similar to the physical world, but differ in some important and pleasant way, and (3) generating complexity. Just as in the physical world, it can produce enjoyment to have significant control or agency in initiating a complex physical interaction, or producing a highly detailed artifact. This could explain the attraction of games such as Minecraft (digital as this example might be), breaking things, or knitting. In the latter example, every aspect of the complexity is controlled, rather than arising from the environment, but it might be argued that some of the satisfaction arises through the same mechanism.

Section Summary: Communication and Collaboration

The implications of the prototypes and lessons presented in this section will be discussed and brought into context in Section 4.4.

107

(a) Multiple users share the virtual space

(b) Painting on each other
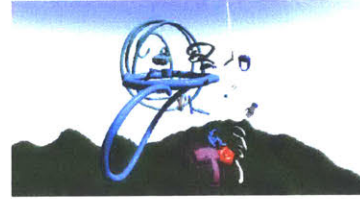
(c) Three users making a sculpture

(d) Setting up a playing field

(e) Playing a game

(f) Playing tic-tac-toe

Figure 4-21: CocoVerse: Collaborative Interaction Examples

4.3 Reality Integration

In Section 3.3, I presented the Reality Integration framework, which concerns methods for integrating virtual reality experiences with the outside world. Issues related to Reality Integration surfaced in some of the prototypes presented in the previous sections; most notably with TerrImmerse (Section 4.1.4) and CocoVerse (Section 4.2.3). In this section, I will first present a final iteration of CocoVerse that addresses issues of Reality Integration that were essential for our industry collaborator to move forward with their real-world application. Then I will present Window, which is a software and hardware tool developed to directly address the cross-contextual issue of spatial integration: allowing local and remote people outside of VR to see into, explore, and interact with the VR space, and vice versa.

4.3.1 CocoVerse Evaluation and Reality Integration

Further development of the CocoVerse project (initially presented in Section 4.2.3) was done in collaboration with Wiley Corning, who did all of the implementation, and contributed to the interaction design.

I traveled to our industry collaborator's site to evaluate the adequacy of our prototype for the intended use case. The basic needs of the use case were met, and the user experience was satisfactory – two users could share an environment inside a 360 photo, insert other photos, and paint. Further refinements were determined to be necessary in order to be fully ready for use. The needs and corresponding solutions are labeled with the relevant category or categories of Reality Integration.

Facilitator needs a method to save scenarios and results. This is a form of social integration, since the facilitator outside VR is a person who fills a contextual role, by being charged with orchestrating the experience of the VR users. It is also a form of spatial integration, since he or she is in the local space with the users. Finally, it is a form of temporal integration, since saving scenarios and results allows them to be used later. To fulfill this need, a save and load system was implemented so that the facilitator could set up the desired scenarios at different locations in a very large virtual space. A special marker object, shown in Figure 4-23(a), was used to add the
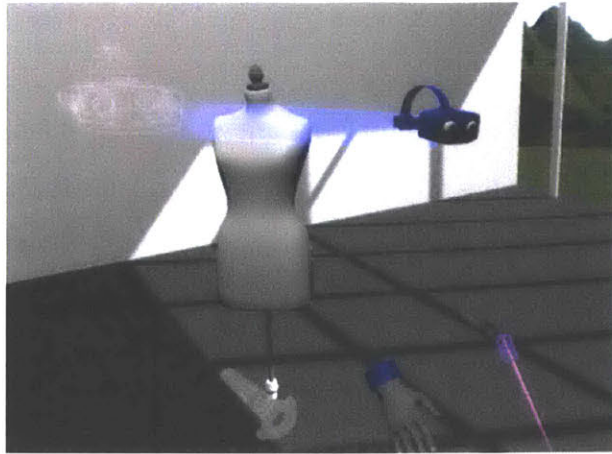
Figure 4-22: Avatar shadows allow users to maintain awareness of other users' physical location, and jump to their virtual location, when desired

desired locations to a list. This object could be instantiated with the 3D printer. The later usage of this saved locations is described in the next paragraph.

Facilitator needs a method to move creators from one scenario to the next, in order. This is primarily a form of social and contextual integration. It is also a form of spatial integration because the facilitator is local to the scenario, and a form of temporal integration, since the scenarios being loaded were previously authored in the VR environment. We implemented a menu on the desktop interface to move creators from one location to the next, shown in Figure 4-23(b). These locations are defined by the marker objects described in the previous paragraph, and are automatically added to the menu that the facilitator can select from.

System needs a method to prevent users from running into each other. This problem fits into social and spatial integration, since it involves establish a proper affordance relationship with people in the local space. In order for users to maintain awareness of each others' physical locations, a wireframe "shadow" stays in place whenever the users' virtual location does not correspond to his physical location. This happens whenever users teleport independently. By pointing the teleporter this shadow, the user can transport to the current virtual location of the corresponding other user. An avatar shadow is shown in Figure 4-22. In this case the displacement between physical and virtual user locations is about one meter.
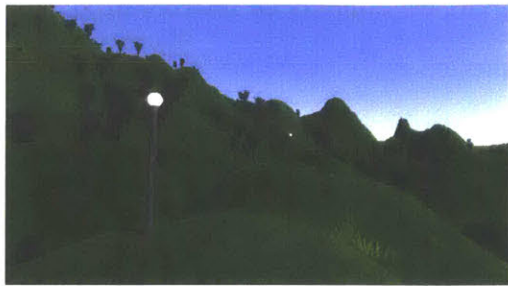
(a) A marker object similar to a street lamp is added to a list of saved locations each time it is instantiated.
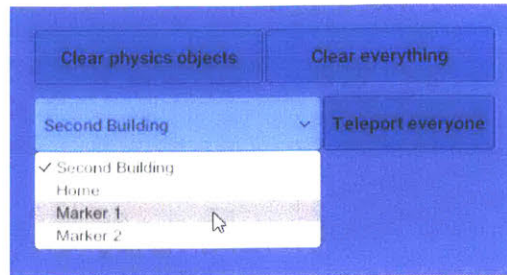
(b) An interface outside VR allows the facilitator to teleport users to the marked locations.

Figure 4-23: CocoVerse: Facilitator Usability Features

Creators need to be able to instantly share their snapshots.   This is a form of temporal integration and social integration, since the snapshots to be shared result from past activity. They can be shared with another person or used within a single users' experience. A run-time syncing system allows host machines to send and receive snapshots taken by each client, and display them when they are placed in the shared space.

CocoVerse: Next Evaluation and Finishing Touches

I traveled to our industry collaborators' site to present our next prototype for further evaluation. This time it was presented to a group of roughly 10 facilitators rather than just three as in the previous visit. With the usage model already fleshed out, this feedback focused on finer details that were nonetheless critical.

- The sizing of objects on selection should be standardized, to avoid sometimes being awkwardly large or small depending on the resolution of the source material.

- A modified architectural space should provide more blank areas to record ideas using sketches, text, and snapshots.

We implemented these improvements and sent them to our industry collaborator. At this time, they have stated that, despite all of the improvements, the user training still requires too much time to be feasible within the actual use case. We hope to do another round of development to address this challenge in the near future. Despite this anticlimactic resting point in our industry collaboration,

111

numerous innovations (bookmark objects, avatar shadows, immediate sharing of 2D snapshots) came out of the requirements of this project and proved both instructive and powerful. Further lessons and conclusions will be discussed at the end of this section (Section 4.3).

4.3.2   Motion Recording for the Equipped Explorer

In this section I present a prototype of a recording tool implemented in the CocoVerse application. The tool can capture and replay actions taken by the user. As such, it allows the Equipped Explorer to explore (in this case, through creation), capture, and review. Using 2D video as a comparison, I will argue that VR recording is a powerful tool for learning, describing its viewing affordances, usefulness as a passive capture tool, and versatility for active content creation.

I use the term *motion recording* (and later just *recording*) to refer to capturing what happens in a VR environment, including the motion of avatars, their interactions with objects, the resultant motion of those objects, sounds associated with these interactions, as well as any other incidental visual, auditory, or other sensory components of an experience that are mediated by the VR system. A motion recording may or may not be complete; that is, it may encompass only some of the elements mentioned. It is broader than *motion capture*, which typically refers only to the capture of the motion of people, and often includes tracking of many points on the body. Motion recording is a form of temporal Reality Integration where interactions are saved for future use. In this section, I distinguish two variants of motion recording. *Passive motion recording* refers to recordings that are not "staged," but are simply recordings of actions and interactions being performed for their own purposes. *Active motion recording* means using motion recording to produce content that is explicitly meant to be viewed by others; this frequently involves repetition and post-processing. These categories may overlap, but the distinction is useful when considering the many ways in which recording can be used.

Viewing and Interacting with Motion Recordings

Unlike viewing 2D video recordings, viewing motion recordings in VR provides access to different perspectives. The recording includes the 3D position and orientation of every object, and therefore every viewing angle is accessible on playback. If a sailor illustrates how to tie a particular knot, the

viewer has no constraints on how close or from what angle she watches. In the case of a virtual lab, multiple parts of a phenomenon being studied may happen simultaneously – through repeated playback, each can be watched from an angle of choice. In the case that the playback takes place in a simulation, the viewer can even change the outcome of the experiment on subsequent viewings by intervening in the scene.

Uses of Passive Motion Recording for Learning

When peers interact in VR, they can resolve their points of confusion. When this scene is recorded, these learners can reinforce their learning later, and other learners can benefit from it as well. Viewing this media is not onerous like staring at a 2D screen – watching a VR playback is much like being there. The voices of participants appear to emanate from the correct physical locations, while head and hand gestures can be viewed in context from different angles.

I argue that VR recording can become a critical part of everyday classroom learning activities, because it creates a rich review experience, and does not require additional time or effort to produce. It can be used as a way to learn (replay what happens when you do X), a way to share with others (look what happened when I did X), and a way to complete classwork (submit a recording of yourself successfully doing X). Prior examples may be preserved and reused by the same instructors, and even distributed more broadly.

Content Creation with Motion Recording

In this section, I discuss the properties of motion recordings from the standpoint of a content creator, who aims to realize a particular dynamic or interactive vision. In the context of learning, this would include students who are creating a piece of content as part of a constructionist learning activity, or a teacher preparing a piece of narrative or interactive content. I will continue to contrast motion recordings with 2D videos, to provide a familiar backdrop for the rather abstract properties being discussed.

The basic principle of content creation using motion recording is to "act it out": for example, if you want to animate a slice of bread flying into a toaster, press "record," grab the toast, and navigate it through the air with your hand into its final resting place. Assuming that the "hand" or other

113

manipulator is hidden upon playback, this method yields dynamic, precise, lifelike motion in a matter of seconds.

2D video is also a performance medium, but there are important differences compared with using VR in an analogous way. Firstly, in the flying toast example above, the simple detail of being able to record the toast without the hand is already a critical difference. Using photo cameras, the same effect must be achieved using hundreds or thousands of individual frames in which the hand enters the scene, moves the object, and then exits the scene again for a still frame to be captured (known as *stop-motion animation*). VR eliminates the need for the "stop" – what's left is *motion animation*. Generalizing this advantage, any recorded object can be individually shown or hidden or shown on replay. Generalizing even further, the entire background scene can be replaced; the individual objects within the scene can be reskinned, duplicated, or overlaid; and time can be dilated or contracted.

When doing traditional video production, the scene may contain heavy physical objects that need to be rearranged. Each time they are rearranged all of the action must be recorded again. In VR, rearranging the set is effortless (nothing physically heavy must be moved), and this can even be done after recording – the analogy would be that everything is done with a green screen, automatically, all the time – and further there are no difficulties associated with placing objects added in post-production at the right size or depth, or difficulties placing such objects at a virtual depth between recorded objects.

One last interesting consideration about VR recording in contrast to video recording relates to the concept of privacy. Distributing 2D videos of individual people interacting with each other or doing an activity carries with it a certain set of privacy concerns– with school children perhaps more than most. VR, on the other hand, sidesteps many of the pitfalls: because the presentation of the human clearly carries its expressivity, but carries its personal·identifiability to a much lesser degree, it should be much more feasible to capture and redistribute interactions between peers and teachers without raising these concerns.

Summing up, content creation using motion recording in VR has advantages associated with its support for powerful kinds of post-processing that are difficult or impossible using 2D video. It also

(a) Activity instructions created in CocoVerse

(b) Learner view while performing activity as instructed

(c) Learner view while reviewing her own performance

Figure 4-24: Capture and Review using the Recorder tool

has the advantage of being highly expressive, while also preserving the privacy of the people being recorded. These properties make it a powerful tool for learning in real-world contexts.

Recording Tool Prototype

To explore motion recording, we added a new tool, the *Recorder*, the explorer's tool belt in CocoVerse. This tool allows the user to record a segment of interaction with the environment– including any other users currently sharing the space and their interactions, and subsequently play it back in place. The playback includes the avatar itself, which allows the explorer to gain a powerful perspective for reflection, which will be explored further– somewhat incidentally– in the experimental study on social presence in Section 5.2.

Figure 4-24 provides a concrete example of the process of capture and review using the Recorder tool. A set of instructions is provided (which the learner is free to follow or not follow), and the learner captures her actions using the Recorder tool. Afterwards, she plays back her own performance, viewing it from a different point of view. As discussed above, this can provide reinforcement and new insights as part of a learning process. Now, if we focus on Figure 4-24(a), we see a complete setup for an interactive activity, including instructions and an apparatus, that were all created in CocoVerse. The objects were placed, their positions fine-tuned, the instructions written, and the location bookmark (recall from Figure 4-23(a)) created using the CocoVerse tools. To accompany it, there is a recorded example of an avatar completing the activity.

The constructionist learning theory places great importance on the creation of interactive artifacts as a part of the learning process . This means that the ability to author interactive experiences

inside the learning environment is significant for two reasons: it empowers teachers to create content without programming, and it provides learners with another avenue for learning.

Reiterating, the example shows (1) how one explorer (learner or educator) can use the tools in the environment to create an activity for another one, and then (2) how that person can learn by completing the activity, and reinforce their learning by using the Recorder tool to review and reflect on that the experience. The creator of the activity uses recording in an active way, as a form of communication, to demonstrate how the activity is to be completed. The participant in the activity uses recording passively for capture, as a way of facilitating and enhancing the process of review and reflection. As such, this prototype illustrates the versatility of the reusable Recording tool to support the Equipped Explorer activities of exploration, capture, review, and communication.

4.3.3   Natural Collaborative Interfaces: Spatial and Social Integration

This work was done collaboratively with Wiley Corning and Dimitri Tskhovrebadze. Wiley did software implementation, and Dimitri built the tracked pen prototype. The gloves were a pre-release commercial product from Manus VR.

In this section I present two explorations of spatial and social integration. The first involves integrating the hands and controllers into VR in a way that more closely mirrors the body and familiar real-world writing tools. The second allows local and remote spatial and social integration by providing a bidirectional spatial window into and out of VR.

Gloves and Pen   One drawback of current VR systems is that the input devices are unnatural – the use holds a wand with buttons that must be learned by sight and touch, but without the visual aid of the hand. Using the Glove and Pen hardware prototypes, this problem is addressed. The result is also that expressive hand gestures involving finger postures can be used naturally. The device prototypes are show in Figure 4-25.

Categories of Reality Integration. The gloves allow users to see other users gestures more vividly, and so this is a form of social integration.

116

(a) Without the Glove, the hand (seen on left), is missing from the virtual environment (right)

(b) The physical Pen (left), is represented in VR using an accurate 3D model (right)
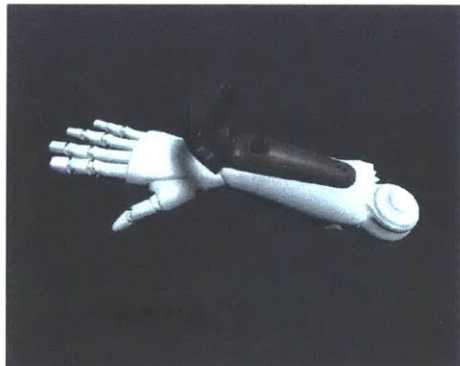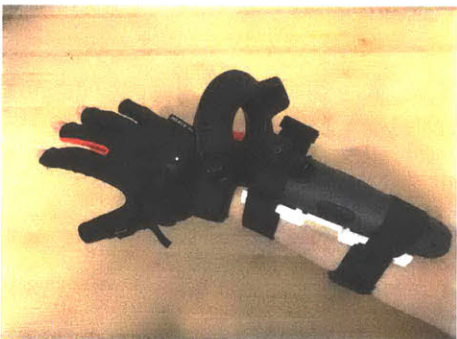
(c) The Glove (left) captures the hand posture so that it can be represented in VR (right)
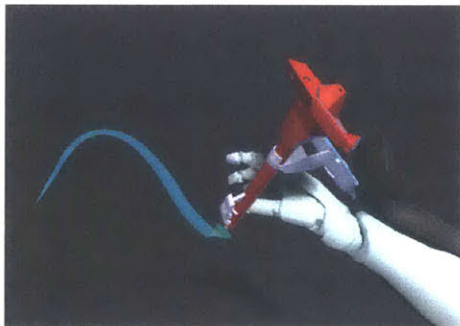
(d) Using the Glove and Pen (left), both the hand and the tool can be represented in VR (right)

Figure 4-25: Natural Interfaces: Glove and Pen

(a) Users outside VR can view the virtual world   (b) VR users can view the real world
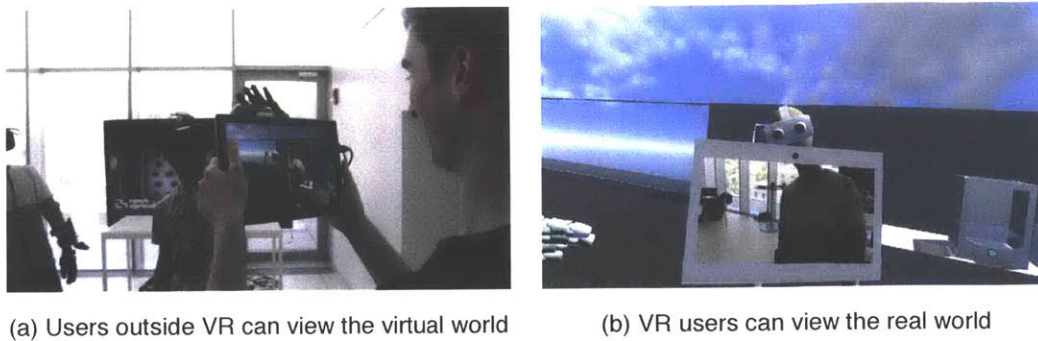
Figure 4-26: Coco Window

Window   One challenge with HMD-based VR is that people in the physical surrounding of a VR interaction cannot easily see what is going on inside the VR environment. The first-person point of view often displayed on a monitor is difficult to follow, and only allows onlookers to look where the viewer is looking. Third-person views generated with a green screen make it easier to understand the context, but they are difficult to set up and don't let the outside user intervene or participate in the VR world in any way. The Window prototype allows users in VR to see and collaborate with users outside VR, and vice versa. This is shown in Figure 4-26.

Categories of Reality Integration. The Window prototype allows local and remote people and spaces to be naturally integrated, and hence falls into the categories of social and spatial integration.

Observations and Lessons Learned   The Glove and Pen system was very effective once calibrated, but did have many limitations, particularly related to the gloves. The orientation of the gloves needed to be calibrated by the user, and the calibration was periodically lost. In addition, while the gloves definitely succeeded at conveying expressive hand gestures for non-verbal communication, there are limitations to the fidelity and accuracy of this representation, which was based on bend sensors for all fingers and an IMU for the thumb. The glove cannot directly infer when the thumb touches another finger, and accordingly this is not properly reflected visually. This attribute feels very unnatural to the user. Overall the project suggested that the dichotomy between the hands and tool or implement (pen in this case) is an effective one.

The Window prototype was proved to be very effective, robust, and practical. Because it uses

the same Vive / SteamVR tracking system as the headset, performance was reliable, and the calibration with the other devices was accurate and dependable. Use of this configuration seems genuinely practical in everyday settings where VR is being used. Such a device could be mounted next to a VR space to show any passersby or colleagues what's going on, and then unholstered to allow the outside person to inspect in more detail or communicate with the person or people in VR. Future work is sure to uncover a plethora of use cases.

Section Summary: Reality Integration Prototypes

Understanding and implementing Reality Integration is sure to be just as important as understanding and implementing in-world affordances such as those described by the Equipped Explorer framework. The Reality Integration related to CocoVerse was implemented for a real-world use case. At first it seemed that the work was mostly done with the first prototype, but we found out through the process of iteration that a great deal was left to do in order to practically enable the application. The use case was ostensibly very straight forward– issues of Reality Integration are sure to be more complex and difficult to solve for more complex use cases. For this reason it is important to focus on generalizability, reusability, and modularity when implementing Reality Integration solutions. Considering the Glove and Pen, and Window prototypes, the lesson was that Reality Integration is powerful. The ideas seemed a cute but not necessarily serious until we implemented them. At that point the significance of all three new forms of connection inside and outside VR became apparent.

4.4 Summary and Discussion of Prototypes

The prototypes and tools presented in this chapter fall into three distinct categories: those that explore learning interactions between the user and the environment; those that explore learning interactions between users; and those that focus on solving practical problems associated with integrating the worlds inside and outside of VR. In each of the corresponding sections, I came away with design approaches, tools, further questions, and future challenges.

In the case of exploration, capture, and review, I came away with the following: a taxonomy for designing simulation-based learning interactions based on fact, fiction, and feasibility; an approach for creating high-bandwidth spatial interactions with simulated systems that appear too rigid to be made interactive; and a user interface approach for capture and review in general use cases. One question remaining at the conclusion of that section was whether there are fundamental advantages to embodied spatial interaction for understanding 3D concepts. This will be addressed by the experimental study presented in Section 5.1.

In the next section on communication and collaboration, I explored multi-user interfaces focused on learning from simulations, and collaborative cocreation. One strong takeaway from all of these was that multi-user interaction in 6DoF VR is highly effective and expressive even with minimal avatars. This claim will be further investigated in the experiment presented in Section 5.2. Beyond that, informally observing a large number of users of a collaborative, open-world cocreation application led to a taxonomy of activities and behaviors that users found highly engaging, related to experiences that are novel, collaborative, or generative. I then drew conclusions about what kind of underlying architecture is likely to give rise to the relevant properties in applications in general. This included the use of tools that are modular and interoperable, so that there is a combinatorial growth in the creation and interaction options available to the user, and a powerful physics engine that supports randomness and dynamics that become increasing complex as the initial states or configurations they operate on become more complex. These important insights will help others to implement environments that are as engaging as CocoVerse but also serve a variety of learning goals.

Finally, I presented prototypes that were directed towards problems of Reality Integration. I showed

the importance of spatial, temporal, social, and contextual integration to enabling practical use cases. Mechanisms like immediate sharing of snapshots in multi-user settings, and saving locations that can be visited in sequence are important and generalizable capabilities that came out of these explorations. The Window prototype shows great promise for bringing VR into real-world settings, by breaking down the barrier between the inside and the outside of the VR world. Finally, the Glove and Pen prototypes explored how the basic interface to the world could be made more natural. The solution was not perfect, but the concept of creating a mapping between physical and virtual hands, along with physical and virtual implements, was also validated.

Summarizing, by using the Equipped Explorer framework as a structure for exploration, I built a variety of prototypes that can act as design exemplars for others. I enumerated generalizable tools and insights for implementing the relevant categories of tools and environments. Finally, I highlighted some solutions to difficulties of Reality Integration, that I hope will help people both to identify and address problems with practically integrating the worlds inside and outside of VR to make VR useful.

In the next chapter I will delve into two specific questions that arose from piloting these prototypes: the possible advantages of using VR to learn about 3D systems (Section 5.1), and the feasibility and possible advantages of collaborating in VR (Section 5.2).

Chapter 5

Experimental Studies

The relevance of the theoretical framework and practical prototypes presented in Chapters 3 and 4 rest on the assumption that it is worthwhile to use VR as a medium for learning. The prototypes stand on their own as evidence that different kinds of applications can be successfully implemented in VR, but they don't answer the question of how VR compares with alternatives. In this chapter, I present two experiments to address this open question in regard to the environmental purposes (exploration, capture, and review), and the social purposes (communication and collaboration) of the Equipped Explorer. The first, presented in Section 5.1, investigates the effectiveness of a learning application in VR by comparing it to a 2D alternative. This question was raised specifically at the end of Section 4.1 as a linchpin for assessing the potential impact of VR for learning. The next experiment, presented in Section 5.2, concerns the quality of communication in VR using minimal avatars. The comparison is made with face-to-face settings, and certain non-obvious advantages are uncovered.

5.1 Learning Differences in VR vs. 2D Using Physics Activities

The work presented in this section has been submitted for publication with coauthors Wiley Corning, Markus Funk, and Pattie Maes. Regarding implementation, this project builds on Electrostatic Playground, presented in Section 4.1.1. The Target Hitting and Field Matching activities were implemented by Wiley Corning, who also contributed significantly to the design. The Field Matching game integrates a 2D electric field visualization implemented by Jonathan Stets with guidance from Professor John Belcher.

As highlighted in earlier sections of this dissertation, VR can provide novel learning experiences–allowing learners to go places they can't go physically, shrink or grow to see the world at different scales, control the passage of time, and hear or see information right when and where it is needed. Related work has found that VR can be beneficial for learning, but the benefit is only seen when the unique attributes of VR are leveraged. The relevant studies have focused on proving out the viability of using VR for certain learning approaches and content, but have not attempted to isolate the specific attributes of the medium that contribute to the benefits obtained. As I hinted at in Section 4.1.1, my goal in this study is to investigate whether experiences that combine dynamic 3D visualizations with spatial sensorimotor interactions provide measurable learning benefits.

Four aspects that define the VR experience are visual and auditory immersion; the presentation of content in stereoscopic 3D; the use of head movement to change perspectives; and direct spatial interaction using hand-based input devices. For the latter two, we refer to movement with six degrees of freedom (6DoF). The first three of these are uncommon for 2D screens, and require the use of additional hardware such as shutter glasses or infrared head tracking. For the last, a common form of direct spatial interaction for 2D screens (albeit 2D rather than 6DoF) uses stylus- or finger-based touch screens.

The goal of this study is to make a side-by-side comparison of the learning that takes place using a standard VR system compared with a standard 2D system. For the best comparison, the interactions and content are made as similar as possible.

5.1.1 Related Work

In Chapter 2, I introduced a variety of background literature and related work highlighting general areas where VR might provide advantages. In this section, I discuss related work that is more closely tied to this specific experiment. In particular, I will survey a body of work concerning learning and training procedures that are highly structured, such as those used in military training. The goal of these studies was to compare VR with 2D interfaces or the real world for spatial memory and procedural knowledge.

A sequence of studies conducted by the US Air Force, published starting in 1992, explored the use of VR simulation-based training. The first pair explored learning navigational and procedural knowledge of small-scale and large-scale spaces, and compared knowledge of real-world spaces with knowledge of virtual spaces acquired during an experiment. Both studies indicated that the use of VR was successful and the knowledge acquired was comparable to that acquired in the real world [RSM92, RY94]. A third study, however, compared the transfer from 2D to the real-world with that from VR to the real world and found no significant difference [Reg97]. The tasks were procedural and navigational. It could be argued that neither took real advantage of the 3D aspect of the environment. The procedural task involved operating a console, which entailed memorizing a sequence of button presses and knob turns. The other involved navigating around a building with two levels– which is technically 3D, but in a rudimentary way. A later study introduced a more sophisticated 3D navigational task– navigating a system of tunnels– and did find that VR was advantageous [SB07].

5.1.2 Interaction Design of Activities for VR and 2D

In this section, I will discuss the design of activities for this experiment in three parts. First I will cover the interaction design– determining what the activities would be, including the nature of the challenge, and the design space for creating a large number of individual exercises. Next, I will talk about specific visual design choices that were made in the process. Finally, I will describe the interface for creating exercises that was used to prepare the experiment.

Activity Subject Matter: Electrostatics

In order to avoid a complex experiment design, we chose activities that help to develop intuition for spatial phenomena, as opposed to remember formulas or solve algebraic problems. With such activities, performance could be quantified (speed and number of attempts), and a multiple-choice test could be based on the activities themselves. In addition, of course, the games needed to be easily adaptable to both VR and 2D interfaces, in such a way as to keep the scale and nature of the interaction highly analogous. We designed two activities that fit these criteria.

Electrostatics is a subject that is introduced in most high school physics classes, and treated in further detail in introductory university physics courses on electricity and magnetism. Some of the basic principles include: there are positive and negative charges; like charges repel, while opposite ones attract (e.g. positive charges repel one another); charges have numerical magnitudes; the strength of attraction or repulsion is proportional to the product of the two charge magnitudes, and inversely proportional to the square of the distance between them; and diagrams using electric field lines can illuminate how a configuration of charges in space will evolve over time [Pur13].

These principles are easy to state, but gaining an intuition for them can be challenging. How does one imagine something that is proportional to the inverse of the square of a distance? What kinds of insights should a particular field line diagram reveal? These are the kinds of intuitions that we made the subject of our activities for the reasons that (1) spatial interaction is directly relevant, since all salient characteristics depend on the positions of charges in 2D or 3D space, and (2) showing comprehension requires only positioning charges in space, so there is no question of irrelevant effects (e.g. changing modes of cognition or interaction to solve an equation or do a numerical evaluation) confounding our results.

Activity I: Target Hitting

The basic premise of the target hitting activity is that a beam of charges must be redirected to intersect a given target. The dynamics of the particle beam follow a physical model of electrostatic interaction (as described above). The participant is given a certain set of charges (sometimes with different positive and negative magnitudes) in order to accomplish this goal. There are particular

126

given positions where the charges may be placed, but the beam trajectory is updated smoothly and continuously as they move particles through space. A diagram of the activity is shown in Figure 5-1.

The reasons to constrain the user to choose from predefined positions– rather than placing charges anywhere in the continuous 3D space– were twofold. The first was to encourage planning and reflection, which we found to be critical to learning. In piloting the activity without predefined positions, we observed that participants would move particles randomly around the space without developing a strategy. Sometimes this led to frustration, while other times they would eventually discover the need to plan and reflect. With predefined positions, it seemed that participants would more quickly discover this approach. After initially moving particles quickly and randomly from position to position, they would realize the combinatorial complexity of the set of possibilities (e.g. there are 42 ways to place two different particles in seven possible locations) and being to plan and reflect. Second, the use of predefined positions supports useful metrics for analysis, looking at the quantity and timing of attempts made to complete the activity.

The design of individual exercises posed an interesting problem that was also solved through iterative design. The relevant parameters are the number of charges to place, and the number of predefined positions given. We initially experimented with a large number of given charges (e.g. five) and a large number of predefined positions (e.g. 20). We observed such activities would take a long time to solve, and induce a great amount of frustration. This also made the time to completion and number of attempts highly subject to individual variation. We also tried introducing "extra" charges that would be left over and not used in the solution, but this exacerbated the problem of difficulty. With these considerations in mind, we settled on exercises with one to four charges and two to ten allowed positions, with all given charges required to complete the activity. Figure 5-2 illustrates these parameters in a typical layout.

Activity II: Field Matching

The field matching activity is similarly based on the idea of moving particles to predefined positions in space in order to accomplish a goal. The goal was to generate a particular configuration of electric field lines through the positioning of the particles, as shown in Figure 5-1. This activity was

127
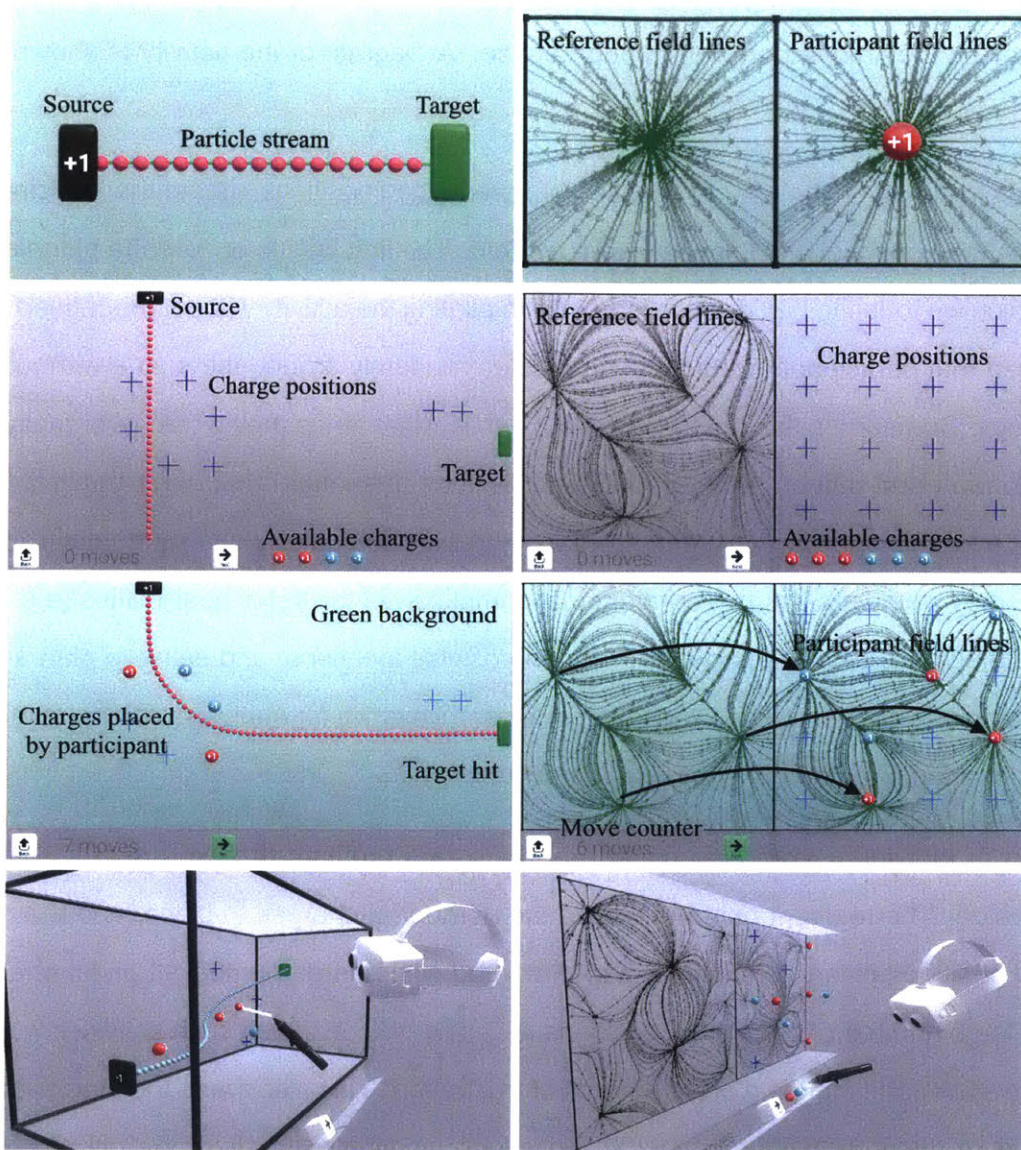
Figure 5-1: Activities: Target Hitting (left) and Field Matching (right). From top to bottom: (i) goal of activity, (ii) initial state of an exercise, (iii) completed state of the exercise, (iv) VR interface
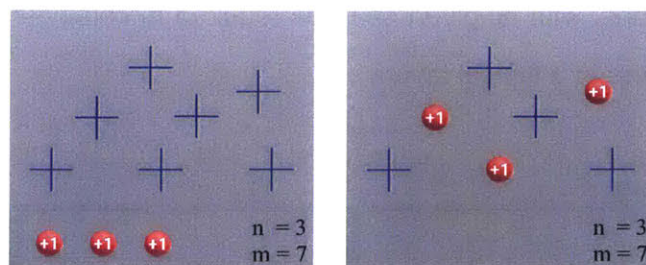
Figure 5-2: Exercise Parameters: n = number of charges to place, m = number of predefined positions. Exercise shown before (left) and after (right) placement.
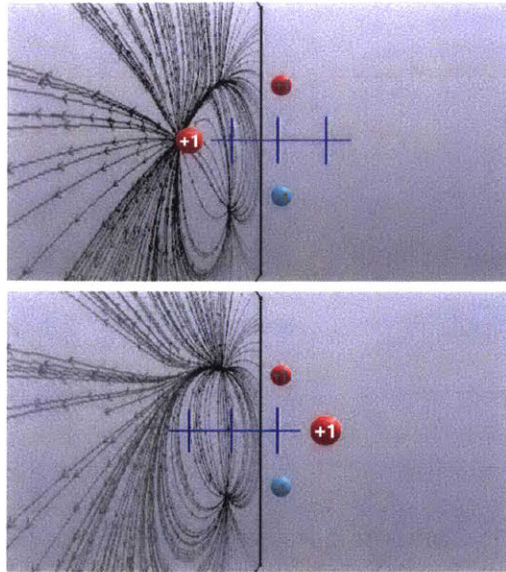
Figure 5-3: Effect of differing z-axis placement in field matching activity. A +1 charge is placed closer (top) or further (bottom) from the field line plane, leading to differences in the pattern of field lines.

designed second, we used predefined positions for the same reasons cited above for the target hitting game.

The set of parameters used in designing exercises for the field matching activity were different from those of the target hitting game. In particular, the field matching activity is generally easier since the field line projections are more likely to reveal (though they do not always do this) where particles must be located. Therefore, the challenge was based more on recognizing attributes of like versus opposite charges that are neighboring, and in more advanced configurations, recognizing the ways in which adjacent particles can lead to a field configuration where it is *not* obvious in which positions particles must be located. Therefore we increased the number of charges the user needed to place, but not as much the number of possible positions, settling on the range of two to nine particles to be placed in four to twelve positions.

Activity Interfaces in VR vs. 2D

Here we describe how the VR and 2D activity interfaces compare, enumerating the attributes of VR and 2D which are the same or different. These are summarized in Table 5.1.

Beginning with the differences, the VR interface supports the following attributes, while the 2D

Attribute	VR	2D
Immersion	Yes	No
Stereo 3D display	Yes	No
Head movement input	Yes	No
Direct manipulation	Yes (6DoF)	Yes (2D)
Primary input	Trigger, move	Pen down, drag
Interaction area/volume	Size of tablet	Size of tablet

Table 5.1: Comparing Activity Interfaces in VR vs. 2D

interface does not: visual and auditory immersion, the use of stereoscopic 3D display, and the use of head movement as an input. Moving on to the similarities, both activities are based on the task of moving charges to positions in space, so we were able to use the same mapping between input methods. The VR interaction requires using the controller's trigger button, while the 2D interaction uses the pen down and pen up actions to initiate and complete charge movement. The VR controller was physically larger, which we were not able to adjust for, but we did make the visual representation analogous. Lastly, we made the interaction space the same physical scale and VR and 2D, with the cross-sectional area of the VR interaction space matching the 2D interaction space.

5.1.3   Visual Design of the Activities

In this section I will briefly visit details of the visual design of the activities. The first regards the appearance of the beam of particles. In the initial design of this activity, the intention was to make the learner set up the initial conditions and then watch in real-time as a single particle would be released from its initial position and then either make its way to the target or not. While this promotes planning, it keeps the pace of exploration very slow. The "beam" visualization was created in order to be able to represent the time dynamics of the system in such a way that the result could be seen instantly. This allows the learner to continuously explore different configurations and instantly see the results of the rearrangements. In order for this to make sense, an increment had to be chosen,
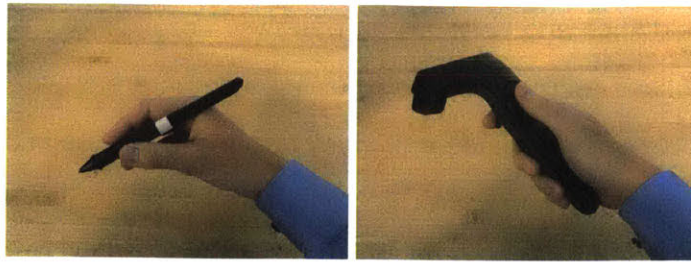
(a) Input devices: 2D (left) and VR (right)
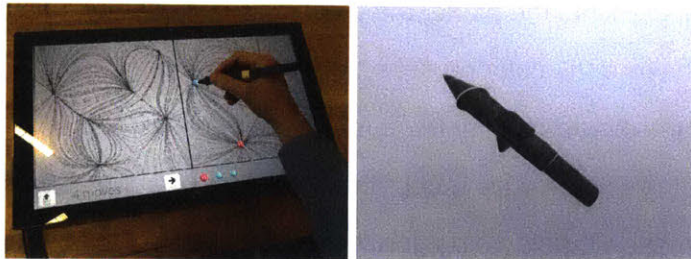
(b) Appearance of input devices in application: 2D (left) and VR (right)

Figure 5-4: Input Affordances for VR and 2D
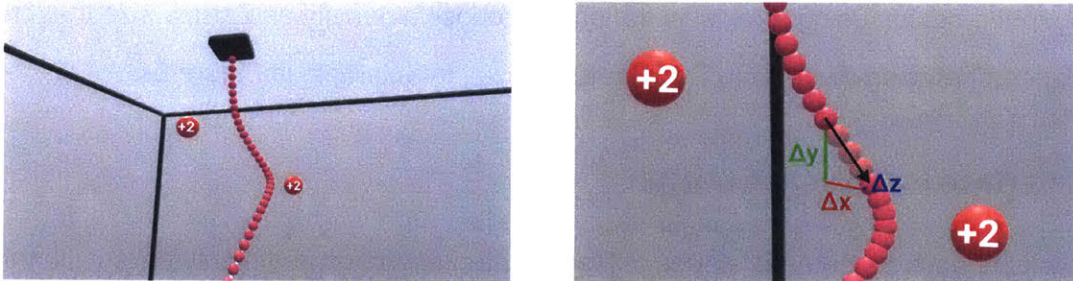
Figure 5-5: Visual Design of Trajectory in Target Hitting Puzzle Game

which would become more sparse when the particle travels faster, and more dense when it travels slower. One possibility was to show this "beam" of future states statically, but this seemed to cause the notion of dynamics and motion to be lost. Therefore, in in the final design all of the individual spaced out particles move in a line, begin spawned at the source, and disappearing at the end of the line. Figure 5-5 shows this visual design in more detail.

In the field matching activity, the biggest challenge in visual design was the representation of field lines. While it is conventional to draw arrows along field lines in static 2D diagrams, this creates limitations in settings where lines are generated dynamically and continuously. In particular, when lines become too close, the arrows begin to overlap and create confusing visual clutter. This would seem to point in the direction of designing an algorithm with many edge cases to dynamically

131

select the density of field lines and positioning of arrows. For example, it could detect the local distance between field lines and refrain from drawing them in areas where the distance is too small. Conversely, it could eliminate lines when they become too dense in certain areas, although this may cause them to become very sparse in other areas.

To address this challenge, we initially took the approach of replacing the use of arrows with the use of color. When projected into the plane, the electric field is defined by a 2D vector at any point. This means that if we have a 2D coloring scheme, we can dispense with the use of arrows. Our idea was to use a unique color to identify the direction of the field, and the brightness to represent the magnitude. In piloting this design, it became clear that, the unconventional nature of it was problematic. The information that is being conveyed is technically clearly distinguishable, but the big issue is consistency– it is inconsistent with other representations of the electric field that are more familiar to learners.

Returning to the design using arrows, in the end we chose a density and arrow size that created a usable tradeoff between clutter and visibility, as shown, for example in Figure 5-3.

5.1.4 Interface for Exercise Design

When conducting an experiment, issues of Reality Integration become highly salient. In this case, it was the creation of activities that required a usable solution. In the course of iterating on the experiment design, we needed to design over 100 different activities. Designing 3D configurations using a 2D interface, or programmatically is awkward and slow. Another problem with program-matically generating activities is that random designs in general do not give the learner a path to discovering the solution. While this might be interesting and useful for advanced learners, it is important to consider the learner's thought process in designing exercises for novices. Therefore we opted to design a VR interface for creating exercises.

Design of Exercise Design Interface

An "exercise" in either Target Hitting or Field Matching consists of (i) a fixed starting configuration, (ii) a set of charges that can be placed in the space, (iii) a set of allowed locations at which to charges can be placed, and (iv) the target location or configuration. The interface we designed is
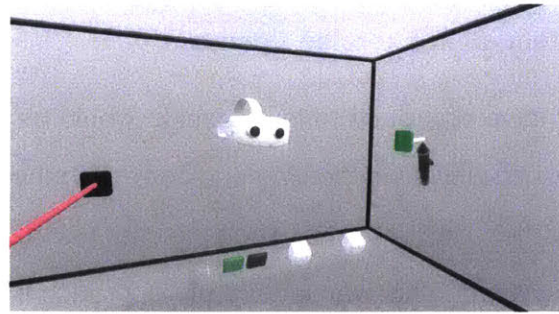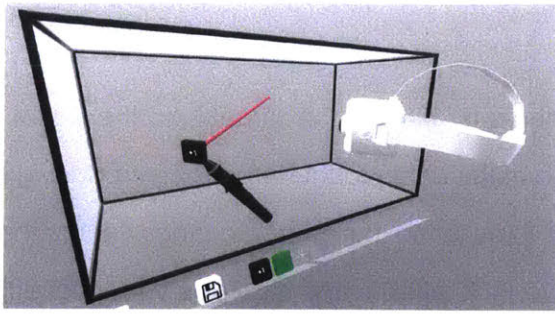
(a) For Target Hitting, the exercise designer can place the beam (left) and target square (right)
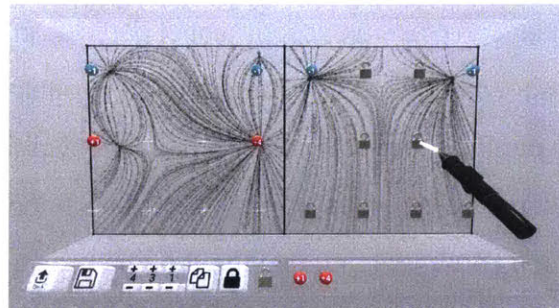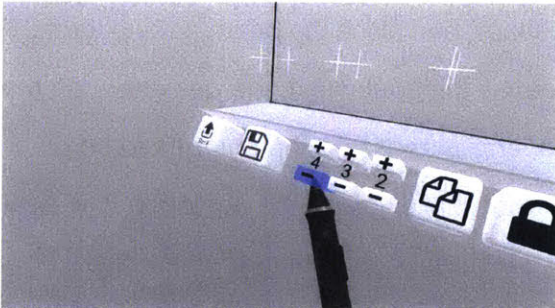
(b) For Field Matching, the exercise designer defines the dimensions of the grid of available locations (left) and then places locks to exclude locations from being presented as options (right)

Figure 5-6: Exercise Design Affordances

very similar to the player interface, except it includes several additional features. Corresponding to each of the four parts of a level mentioned above are affordances in the level design user interface. To accomplish both (i) and (ii), the designer has the ability to create charges. This way, both the charges in the environment, and the charges in the game tray can just be placed there. For (iii), the designer is given the ability to place the allowed sites. These are marked with white "axes," and the designer can place these, and experiment by placing player charges at different locations, to understand the possible outcomes of the game. In the case of Field Matching, we made the choice to fix options on a grid. This simplified the design, and made it easier to deal with the insensitivity problem introduced above. Spacing the grid out parametrically allows the designer to work within a set of options where the differences between choices are guaranteed to be visibly large. Finally, defining the target configuration (iv) works slightly differently for Target Hitting versus Field Matching. In the case of Target Hitting, green square targets are just another object that the designer can place. In the case of Field Matching, the target is derived directly from a set of charges that are then "hidden" when the target pattern is presented to the learner.

Challenges in Exercise Design

One naïve approach to level design would be to begin by placing the target and the beam, and then successively deflecting and correcting the path of the beam to the target by introducing pairs of charges, one at a time. Then the charges that are to be left to the learner to place can be the last ones placed. This was indeed the approach that we took initially, and it proves very challenging. An alternate approach that makes designing exercises very easy is as follows: in any order, the designer can place an assortment of particles in the space, and place the beam among them. Then, the target can be set anywhere along the beam to define the goal. Next, the designer can remove a few particles– any subset– from that configuration, and let these be the charges to be placed by the user. This provides a linear process with no inherent challenges to the designer. The designer can focus on understanding the thought process of a learner approaching the exercise.

Exercise Overview: Incidental Interface Affordance for Real-World Deployment

In Section 2.2.5, I discussed note-taking and the importance of memory cues. As learners progress through sequences of levels, they accumulate key insights into unsuccessful and successful solution strategies. The ability to review these can help to solidify and integrate these learnings. Visual and spatial overviews allow people to cue their memory simultaneously for many different ideas or contexts. Making overviews in VR can leverage simultaneous memory cuing, just like in 2D, but it has different properties. In particular, an immersive overview cannot be seen all at once, but one can be cognitively aware of the entire space after inspecting it. Furthermore, the number of viewing perspectives is large, so that anytime the scene or context is viewed, new insights may be revealed.

Incidental to the implementation of the design interface was our implementation of such an overview mode. Learners (and designers) can see completed levels in one direction and upcoming levels in another, using scaled-down representations arranged along a linear timeline. This is shown in Figure 5-7. A question for further research is what effect this might have on learning when it is used delibrately – whether it is better because memory is cued and new information provided at the same time, or worse because the restoring of context is hindered through the introduction of variations.
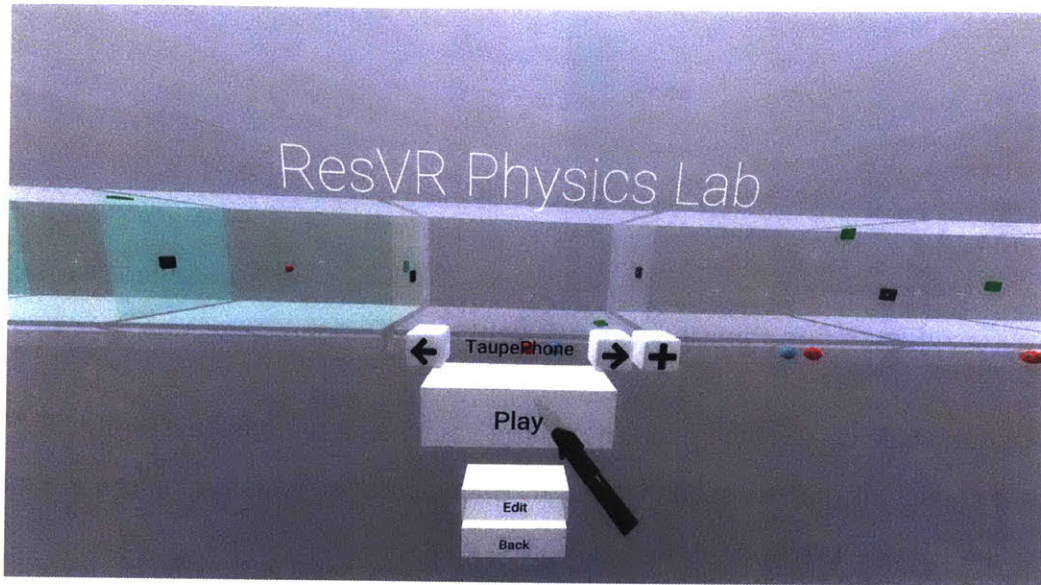
Figure 5-7: Exercise Timeline Overview

5.1.5 Experiment Comparing Learning in VR and 2D

We sought to compare the effect of 2D versus VR interaction on learning. In our experiment design, each participant completes each of the activities, one in 2D and the other in VR. We counterbalanced the order in which the VR and 2D activities were performed. A two-part multiple choice pre/post test assesses participants' competencies associated with each of the activities.

Method

Hypothesis The hypothesis was that learners would do better on the multiple choice test and complete activities faster and with fewer moves when trained in VR, both (i) immediately, and (ii) after two weeks.

Independent Variables The independent variables in the experiment design are the modality (2D or VR) and the activity (target-hitting or field matching).

Dependent Variables In order to measure the effect of the independent variables, we measured the following dependent variables:

- Completion times and number of moves required during testing activities, during first and second sessions.

135
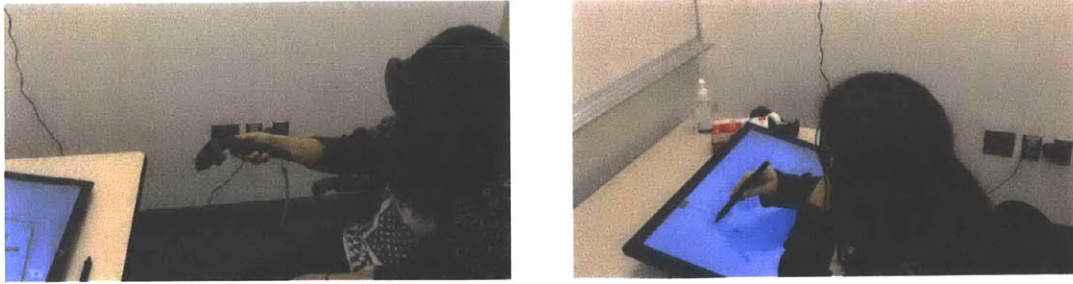
Figure 5-8: Participants in VR (left) and 2D (right) conditions

- Two-part multiple choice test, performed three times: as a pre/post test for the first session, and at the beginning of the second session. Each of the two separate parts were relevant to one of the activities. The identical test is given all three times.

- TLX (perceived cognitive load) was measured for every 2D or VR activity session.

- Text-based questionnaire comparing activities and 2D/VR

Apparatus

The 2D activities are completed on a 22-in drawing tablet (Monoprice 114481). Moves were performed using a touch-and-drag interaction with the stylus. The VR activities run on the HTC Vive and use the controller's trigger for clicking and dragging charges in space. Participants are shown in the two conditions in Figure 5-8, and the 2D and VR designs for the two activities are shown in Figures 5-1 and 5-1.

Multiple Choice Test The multiple choice test involved static graphical representations of systems that matched the dynamic interactive activities used in style and content. One example is shown in Figure 5-9. Half of the questions were relevant to the target hitting game, and the other half were relevant to the field matching game. These questions were appropriate for use in a pre-test, since the basic representations (charged particles, electric fields) are standard.

Participants

We invited 20 participants to take part this experiment, in the age range 18-22. They had all taken MIT's electricity and magnetism course in the past three years, so they had a familiarity with the subject matter of the activities. Participants were split into 4 groups according to the

136
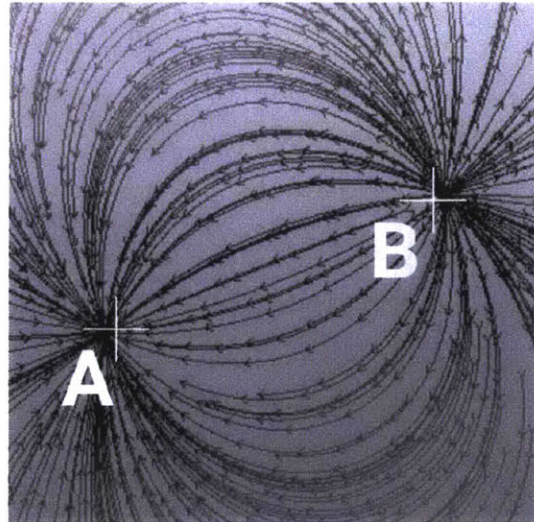
Based on the scenario in the above picture,
Which configuration of charges would produce the field lines shown?

POSITIVE charge in Position A, NEGATIVE charge in Position B

NEGATIVE charge in Position A, POSITIVE charge in Position B

POSITIVE charges in both positions

NEGATIVE charges in both positions

Figure 5-9: Example Multiple Choice Question

counterbalanced orders of the activities and modalities. Due to technical difficulties, we had to exclude two of the 20 participants.

Procedure

Participants were required to attend two sessions, separated by 12 to 16 days. The sequence of activities in the first session was as follows: they were given the multiple-choice pretest, completed each of the two activities in the modality appropriate to their group with training (modality varied) and testing (always 2D) portions, completing the TLX questionnaire after each testing segment, and finally they were given the multiple-choice post-test (identical to the pre-test). Participants were told they were being timed at the beginning of the overall session, but they were not explicitly reminded between activities.

In the second session, the sequence of activities was as follows: participants completed the multiple-choice test, and testing sets for each game, completing a TLX questionnaire directly after
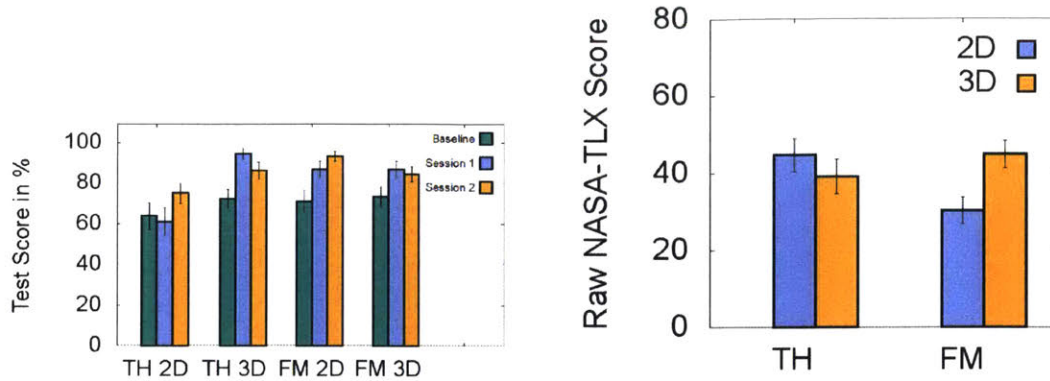
Figure 5-10: Test scores and perceived cognitive load (TLX) for Target Hitting (TH) and Field Matching (FM) activities
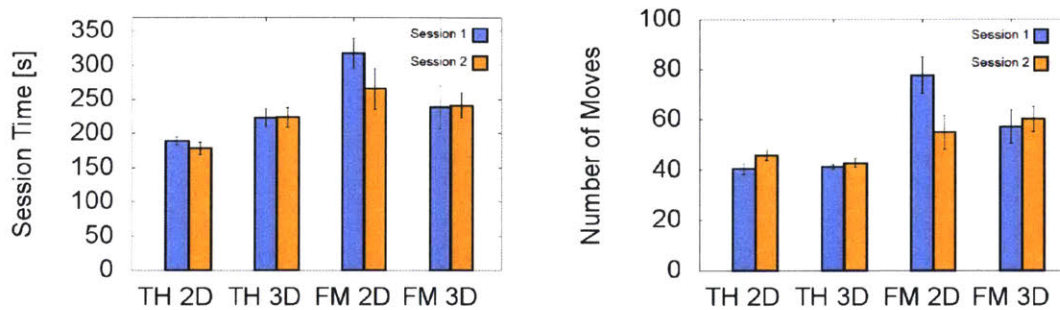
Figure 5-11: Times and attempted moves for Target Hitting (TH) and Field Matching (FM) activities

each. At this point, all of the tasks relevant to our quantitative analysis were complete, and some additional tasks were performed to gather further qualitative feedback. In particular, we wanted participants to be able to contrast the two activities and the two modalities. Recall that they had each tried only one of the two activities in VR during the earlier portion of the study. Therefore, participants were next given exercises for both activities in VR (one of which they had not seen before). These exercise sequences were shorter than the the ones during the earlier part of the study, since the idea was for them to gain experience with both activities in VR– not to assess their performance, as before.

Finally, they were given a free-response, text-based questionnaire about the differences between the activities and the VR versus 2D experiences. The complete list of questions is given in Table 5.2.

138

Results

Quantitative Performance Results For statistically comparing the session times, the number of moves, the results of the multiple choice questions, and the NASA-TLX score, we are using a two-way ANOVA. Levene's test could not find a difference in the error variances (p>.05). We were using a Bonferroni correction for all post-hoc tests.

Target Hitting First, we analyzed the performance of the participants considering the multiple choice questions in the target hitting game. For the 2D modality, the participants scored 63.89% (SD=28.26%) for the baseline, 61.11% (SD=28.26%) for after the first game, and 75.00% (SD=21.65%) after two weeks. Considering the VR modality, the participants scored 72.22% (SD=19.54%) for the baseline, 94.44% (SD=11.24%) for after the first game, and 86.11% (SD=18.16%) after two weeks. A two-way ANOVA found a significant effect in modality, $F(1, 48) = 8.647$, $p = .005$. The effect size estimate shows a large effect ($\eta^2 = .153$).

We compared the session times between the two sessions for the target hitting game. For the 2D modality, the times after the first session (M=189.56s, SD=18.80s) were higher than the times of the second session two weeks after (M=178.44s, SD=26.06s). Considering the VR modality, the participants took an average of 223.47s (SD=40.37s) for the first session and an average of 223.80s (SD=43.76) after two weeks. A two-way ANOVA found a significant effect in modality, $F(1, 32) = 12.357$, $p < .001$. The effect size estimate shows a large effect ($\eta^2 = .279$).

Further, we compared the number of moves that the participants made in the activities between the two sessions for the target hitting game. For the 2D modality, the number of moves in the first session (M=40.44, SD=5.91) were fewer than in the second session two weeks after (M=45.89, SD=6.12). Considering the VR modality, the participants made slightly fewer moves in the first session (M=41.33, SD=2.44) than in the session after two weeks (M=42.67, SD=5.33). A two-way ANOVA did not reveal a significant effect in modality or session.

Finally, we compared the perceived cognitive load between the VR and the 2D variant of the target hitting activity using the raw NASA-TLX score. The 2D variant of the training activity (M=44.56, SD=12.79) was perceived as a little more cognitively demanding compared to the VR variant

(M=39.11, SD=13.57) of the game. However, a one-way ANOVA test could not reveal a significant difference.

Field Matching Also for the field matching game, we analyzed the participants performance according to the modality and the trial. For the 2D modality, the participants scored 71.11% (SD=22.60%) for the baseline, 86.67% (SD=17.32%) for after the first activity, and 93.33% (SD=10.00%) after two weeks. Considering the VR modality, the participants scored 73.33% (SD=20.00%) for the baseline, 86.67% (SD=17.32%) for after the first activity, and 84.44% (SD=16.17%) after two weeks. A two-way ANOVA found a significant effect in trial, $F(2, 48) = 4.682$, $p = .014$. The effect size estimate shows a large effect ($\eta^2 = .163$). The post-hoc test revealed a significant effect between the baseline and the two weeks after trial for the field matching activity.

We compared the session times between the two sessions for the field matching activity. For the 2D modality, the times after the first session (M=317.21, SD=66.82) were higher than in the second session two weeks after (M=265.57s, SD=90.91s). Considering the VR modality, the participants were faster in the first session (M=238.62, SD=92.53s) than in the session after two weeks (M=241.46, SD=54.72). A two-way ANOVA did not reveal a significant effect in modality or session.

Further, we compared the number of moves that the participants made in the activities between the two sessions for the field matching activity. For the 2D modality, the number of moves in the first session (M=77.44, SD=21.57) were higher than in the second session two weeks after (M=54.89, SD=19.98). Considering the VR modality, the participants made slightly fewer moves in the first session (M=57.22, SD=20.02) than in the session after two weeks (M=60.22, SD=14.98). A two-way ANOVA did not reveal a significant effect in modality or session.

Finally, we compared the perceived cognitive load using the raw NASA-TLX score for the field matching activity. The 2D variant of the activity (M=30.22, SD=10.12) was perceived as less cognitively demanding compared to the VR variant of the activity (M=44.78, SD=10.50). A one way ANOVA test revealed a significant difference between the 2D and VR variants, $F(1, 8) = 17.135$, $p = .003$. The effect size estimate shows a large effect ($\eta^2 = .682$).

Free Response Questionnaire Results   Three of the free-response questions were concerned with the comparison between the two activities, while the other three were concerned with the comparison between 2D and VR interaction. These sets are discussed first separately, and then brought into context in an overall discussion. The text of the questions is shown in Table 5.2.

Comparing Target Hitting with Field Matching   The three free-response questions concerning the comparison between activities inquired respectively about comparative (Q1) difficulty (which was easier), (Q2) amount of learning, (Q3) engagement. Since these questions asked the participant to choose one activity or the other, the responses could be coded according to their choices, and additional nuance obtained from their elaborations. Table 5.3 shows participants' coded responses.

There was a complete consensus among participants that the field matching activity was the easier of the two (Q1). There was a mixed response about whether they learned more from either activity (Q2), or equally from both. The largest group said that they learned more from the harder activity, but this group constituted less than half of the respondents. The questions about engagement (Q3) exposed participants' ambivalence about being challenged. Specifically, 72% of respondents said that the target hitting activity was more engaging, but 33% explicitly (6 participants) mentioned the tension between feeling satisfaction while doing well at the easier activity, and the rewarding but also frustrating experience of doing something more novel and difficult while doing the harder activity.

Comparing the 2D and VR experiences   The other three of the free-response questions aimed to uncover differences between the 2D and VR experiences. Unlike the first three, these were formulated as "how" questions (see Table 5.2), so the coding scheme was created manually post hoc. Most responses included multiple phrases, and were correspondingly assigned to multiple codes. The codes and results are shown in Table 5.4.

In this section, we will take a factual approach to reporting the results, and then discuss the implications in the next section. I would first like to comment on how these results should be technically interpreted. The key point is that the coding scheme was applied to spontaneous responses to broad questions. For example, one third of respondents mentioned that the *VR activity was chal-*

Number	Category	Question
Q1	Activities	Was one activity easier than the other? If so, why?
Q2	Activities	Do you feel like you learned more from one activity than the other? If so, why?
Q3	Activities	Was one activity more engaging or fun than the other? If so, why?
Q4	2D/VR	How did your experience doing TH (the activity with the particle stream) differ between the 2D version and the VR version?
Q5	2D/VR	How did your experience doing FLM (the activity with the field lines) differ between the 2D version and the VR version?
Q6	2D/VR	In general, how did the VR interface compare to the 2D interface?

Table 5.2: Free Response Questions from Second Session

Answer	Q1 (easier)	Q2 (learned more)	Q3 (more engaging)
TH	0	8	13
FM	18	4	2
Neither/same	0	6	3

Table 5.3: Free Response Results for Questions Comparing the Two Activities

Code	Description	Q4 (TH)	Q5 (FM)	Q6 (interface)	total
1	Novel spatial insights in VR	16	6	8	32
2	VR cool, fun, engaging, interesting	11	7	13	31
3	VR activities challenging or harder	6	8	0	14
4	VR/2D experiences similar	2	7	5	14
5	VR had disadvantages	0	9	3	12
6	VR interface was better	2	0	5	7
7	2D interface was good	0	4	1	5
8	2D interface was hard	3	1	0	4

Table 5.4: Free Response Results by Category for Questions Comparing VR with 2D

lenging or harder (Code 3) in response to *Q4: How did your experience doing TH differ between the 2D version and the VR version?*. The fact that six different people spontaneously made the same point provides strong evidence in favor of the point– clearly very different from something like "one third of respondents agreed with the statement."

To begin, we'll discuss *Q4*, which concerned the 2D vs VR experience of the Target Hitting activity. The *Code 1* rate of 88% indicates that all but two participants reported novel spatial insights in VR. Some examples are: "The VR version's 3D aspect made the game more interesting, and also made the game feel more representative of a physical scenario." and "Having the particles streams in space allowed me to better visual [sic] the effect the added particles would have on them." Next, 62% of respondents reported *Code 2*, that VR was fun, engaging, or interest. Some example statements: "It looked super realistic and was more fun to play and experiment around with." and "It was more interactive which made it more fun." [referring to VR].

The statements for *Code 3* (33% of participants) indicated that the VR activity was more challenging. Respondents referred to the inherent additional complexity of solving a 3D challenge compared with a 2D puzzle: "Adding a 3D component made in [sic] much more challenging." and "The VR version is definitely harder because you have another dimension to take into account." However, 16.7% of participants made a contrasting statement that the 2D version was harder. A closer look at these responses reveals that these participants were referring to the difficulty of the specific exercises they were presented with, as opposed to the inherent properties of the activity: for example "I think some of the 2D puzzles were either harder or harder to visualize." This contrast was made explicit in a single response which made both points "The 2D version I thought was harder than the VR version ... I thought the VR version was more fun because there was a 3rd dimension that made it a little trickier." The questionnaire was administered in Part 2 of the experiment (the second session, two weeks after the first), where the VR sessions were actually shorter and easier than the 2D versions. These sessions were included after the conclusion of the activities measuring performance, to give the participants a taste of both activities in VR, one of which they had not done in VR previously (as dictated by their experimental group).

Next we'll move on to *Q5*, which concerned the contrast between Field Line Matching in VR versus

2D. Participants were not as enthusiastic about the value of this activity in VR. One third reported novel spatial insights— comparatively fewer than for Target Hitting. Two attributes of the field-line activity seem to account for this: firstly, the activity was seen as being similar between 2D and VR versions, in particular because the field line plane was in 2D in both 2D and VR activities. Furthermore, the remaining 3D aspect— that the field lines in the field line plane result from charges outside the plane— was considered interesting by some, but confusing by others. The former contingent made this statements: "I was really surprised because even after learning physics E&M in both high school and at MIT, it never occurred to me that electron FLM vary in 3D, and it was just really interesting to realize." "the 3D part of the VR version seems quite useful for getting an intuition of how field lines look in 3D."), while the latter made these: "the building of a 2 dimensional field line diagram in 3d space is a little confusing." "The VR version was confusing at first." "we still only had a 2D projection of the field lines, which was weird.". Multiple participants highlighted that it was difficult to distinguish the influence of depth (Code 5): "It was hard to figure out how far the charges are affected the 2D drawing." and "The effect of the charges on field lines in space was less clear." [referring to VR version], and a third participant went further described how this led to a trial-and-error strategy "it was tough too distinguish between the field lines when a charge is placed in two similar locations (that have a different depth). Because of this it required some trial and error."

Finally, in *Q6*, we asked participants to reflect on the general difference between VR and 2D, with a focus on the interface. The most common response focused on the fun, engaging, and interesting aspects of the VR experience. "It was more mentally stimulating and more fun. Whereas I got frustrated more quickly doing the 2D puzzles when I couldn't solve something, I was entertained when I couldn't solve the VR puzzles." "VR made the games more interesting and interactive, which made them more engaging."

Comments on the advantages of the VR interface (Code 6) described it being easy, smooth, or natural: "The VR interface felt a lot more natural and engaging than the 2d one. I liked being able to place the charges in a real way that made it easier for me to understand what my placement would do.""Overall I like the VR interface better than the 2D interface. It is easier to navigate and move around the particles." "I think it was also easier to move particles around in the VR version

than 2D version.", "The VR interface felt very smooth and elegant." An equal number of comments indicated that the 2D and VR experiences were similar (Code 4): "Very similar interfaces. VR seemed sort of unnecessary - it was cool to have, but ultimately the same goal could be reached on the 2d interface.", "The VR interface and 2D interface were about the same in terms of being an intuitive / easy to use interface.". Two participants distinguished explicitly between the activities, making both points, each in reference to one of the activities: "I really enjoyed the VR interface for the particle stream game. I think having that extra dimension is really exciting and adds another way that the particles you place can mess up your streams. The field line matching one felt the same regardless of interface, but it'd be interesting to see more use of the fact that's in 3D otherwise I feel that the 2D interface accomplishes what the 3D one does in its current iteration without the need for the fancy setup." "The VR interface allowed for added complexity with respect to how charges interact. It helped with rounding out understanding for the beam game [target hitting], but may have added confusion to the FLM game."

Discussion

There were some expected results and some unexpected. Considering all of the data, we are able to establish a clear picture that explains most of the unexpected results.

For the target-hitting activity, participants who trained in VR performed significantly better on the multiple choice test: the VR group moved from 72% to 94% in their first session, while the 2D group stayed the same, scoring 64% and then 61%, as shown in Figure 5-10. At the same time, the VR group took significantly longer to complete the activities, while the number of moves tried did not differ significantly. This implies a greater time between moves on average, painting a picture of a more pensive approach. To reiterate two important aspects of the experimental setup, the completion times and number of moves we're referring to here correspond to participants' behavior using the same modality (2D) to complete the same set of activities. That is, the training phase takes place in different modalities, while the testing phase is identical in every way across all groups. Also, there was no difference in how or when they were advised they were being timed or the importance of completing the activities quickly. To summarize, when having trained in VR, participants approached the 2D testing activity more pensively, and in the end performed

145

significantly better on the multiple choice test.

Participants did not exhibit the same learning benefit for VR in the field matching activity. The qualitative feedback clarifies the reasons for the this. In particular, the skill being trained (arranging charges to create a pattern of field lines) seems intuitive in 2D, but confusing and unfamiliar in VR. Participants mentioned never been introduced in their academic coursework to the concept of projecting field lines onto a plane. In hindsight, while it is a valid concept, it is not clear when or how it would be used practically. It could have been motivated by the design of the activity (e.g. trying to influence a particle constrained between two surfaces), but it was not. In addition to that shortcoming, participants reported that the effect of the z-axis was too hard to discern, adding frustration to lack of motivation. It is worth noting that despite this, their approach in the testing phase was not significantly different, nor was their performance on the multiple choice test.

Now, taking a step back, we have the noteworthy result that doing an activity in VR was able to better prepare learners for a 2D test than was practicing the activity in a 2D interface much closer to the test. That is, learners were able to overcome the barrier of skill transfer, which impedes the application of knowledge acquired in one setting from being applied in another setting, and perform significantly better. Correlated with that better performance was a measurable difference in approach– the VR-trained learners took more time when completing 2D activities. So there are at least two different effects here: the first was the advantage of engaging with a 3D system in a visual and sensorimotor context , and the other was taking a more pensive and reflective approach to the learning process.

The difference in learning approach was not something that we predicted prior to the study– on the contrary, we had hypothesized that the VR learners would be faster and use less moves. In hindsight, however it seems clear that the learning assessed by the multiple choice test at the end need not correlate with the speed of completing the 2D activities. A general takeaway here is that behavioral metrics in completing exercises can in fact give insight into the learning that is taking place, but naïve interpretations might lead to exactly the wrong conclusions. Had we not included the multiple choice assessment in our study, we could have concluded that the VR inter-face had been inferior at preparing students for the 2D activity; since they took longer to complete

the exercises.

5.1.6 Conclusion and Outlook

To conclude the discussion, we remark that the VR learning advantage we've demonstrated here may be the tip of a very large iceberg. Of course we are not the first to suppose that learning in VR might have advantages, as noted in the *Related Work*, but we are hoping that there will soon be a much larger volume of studies on learning in VR that refine both our knowledge and our research methodologies. It is remarkably difficult to evaluate the efficacy of any aspect of a learning environment, given the vast individual differences one finds even within a single classroom. VR does offer some advantages here, in that it facilitates collecting behavioral metrics during learning activities. In this study we recruited a group of students who had all taken a particular course, and gave them a set of learning activities that constituted less than half an hour of learning. We constrained the use of immersion, 3D, head, and hand-input in a way that could be easily mapped to a 2D analog, and still found a significant advantage. It is easy to imagine that the comparative benefits obtained through the use of VR over longer periods and using a more comprehensive and versatile learning environment might be much greater still. This appears to be a fruitful and exciting area for future work.

The most salient questions, moving forward, fall into three categories: First is identifying and proving out the fundamental advantages of the VR medium. One example is researching the consequences of linking sensorimotor activities directly with dynamic simulations, the area where the current work contributes. There may be advantages to accessing existing materials in a way that is more intuitive, eliminates distractions, or decreases extraneous cognitive load [RSM92]. Second, is exploring how to apply existing instructional strategies, or inventing new ones, that uniquely leverage the combination of deep similarity to the real world (human-scale or otherwise) and the great freedom to design information presentation afforded by VR [DSL96, ROS06]. This includes learner pacing, smart tutoring that is contingent on attention [HSZ17], merging spatial representations [RMF15b, Ain06], and facilitating novel forms of collaboration with peers and experts [KB15]. The latter is the topic of the experiment presented in, the next section (Section 5.2). Third, the combination of the two previous categories should give rise to design guidelines and specific

knowledge about subject areas and usage scenarios where the advantages of VR are the greatest [Pan10, SC13]. Identifying these is a category of its own, that lies closest to direct real-world applicability. In fact, this is the area where the industry should be encouraged most to participate in the research community, allowing the latter to benefit from their often greater ability to do large-scale testing, evaluation, and hence new insights into design requirements. In turn, the research community can stay focused on relevant problems, and contribute applicable knowledge.

These conclusions will be brought into context with conclusions from other parts of the dissertation in Chapter 6.

5.2 Social Presence and Communication with Embodied Avatars

The work presented in this section was the basis for a publication I co-authored with Zhangyuan Wang, Markus Funk, and Pattie Maes [GWFM17]. I would also like to thank Hisham Bedri and Ronen Zilberman for their contributions to the project, and Chris Schmandt for his useful feedback.

In Section 3.1, I enumerated categories of activities and tools that are important for learning in VR, including communication and collaboration as one of them. Some prototypes of multi-user experiences were presented in Section 4.2. Informal trials indicated that users sharing the virtual space felt a strong sense of social presence, despite the minimal nature of the avatars we used. This seems to provide a strong foundation for communicative and collaborative learning experiences in VR, and the goal of the study presented in this section is to delve deeper into this phenomenon and gain a more specific understanding of the mechanisms involved.

Embodied room-scale virtual reality endows users with a very different relationship to their own avatars and virtual environments than analogous non-immersive systems, which use input from keyboards, mice, gamepads, and joysticks. Users *embody* their avatars in a direct way – movements are one-to-one at physical scale, and they move and reach naturally in order to interact with objects. One dual implication of this fact is that when observing others' avatars in the virtual environment, they *look* human. That is, the same precise measurement of movement that is required to deliver the first-person VR experience allows these movements to be made visible to others as body movements with great fidelity. Consequently, when two people share a virtual space in this fashion, they each have a strong sense of being present with another human. Prior works have established the general principle that high *movement realism* achieves a strong sense of social presence, using comparatively low information-bandwidth.

The system, an extension of CocoPaint (see Section 4.2.2), allows two users to interact in room-scale VR (i.e. six degree-of-freedom tracking of head and two handheld controllers) in the same-time, same-place setting, and is the first of its kind that we are aware of in the research community. The goals of this experiment are to (1) establish the basic feasibility and utility of this kind of multi-user interaction, (2) pilot methodologies for studying behavior in this setting, (3) offer early results

Figure 5-12: Same-time, same-place inter-
action in room-scale VR

Figure 5-13: Avatars for Charades and Pictionary

related to similarities, differences, advantages and disadvantages compared with face-to-face, (4) explore the use of freehand drawing in 3D for communicative interaction, and (5) propose future research directions. The choice to use minimal avatars was made to avoid complicating our results with effects related to the choice of body representation– some such effects will be mentioned in the related work below.

We designed a set of goal-oriented, communicative activities for pairs of participants to perform in an experimental setting. These were popular word-guessing games based on gesturing and freehand drawing that could be directly compared in face-to-face and VR settings. The different words that participants attempted to communicate represented a broad array of concepts and corresponding symbolic gestures. These are considered to be proxies for various communicative face-to-face activities. To explore the use of 3D drawing in communicative interaction, we had participants play an analogous game using freehand drawing in 3D instead of 2D. We evaluated the experiences using a combination of methods and metrics: the VR system itself provided data on movement; electrodermal activity was captured to measure engagement; users completed questionnaires measuring perceived mental load, presence, and other aspects of the experience; participants did think-aloud reflection while reviewing recordings; and semi-structured interviews were conducted with participants to gain further qualitative insights.

In the sections that follow, I first discuss related work, then briefly describe the system (similar to that used for CocoPaint in Section 4.2.2) that we built for collaboration in room-scale virtual reality in the same-time, same-place setting. Next I discuss the experiment we carried out, which required extensive modification and adaptation of the basic system, and present the corresponding results.

150

Then I discuss the implications of the quantitative and qualitative results of the experiment. Finally, I conclude and highlight promising directions for future research.

5.2.1 Related Work

In Chapter 2, I presented background literature that points to potential advantages of learning in virtual reality. The question of the effectiveness of communication in 6DoF VR– what factors are necessary to make it work, and what differences or possible advantages are there compared with face-to-face– arose in the context of the prototypes presented in Section 4.2. For this reason, there is additional background to introduce in the context of this experiment. Two related bodies of research I'll discuss focus on (i) the psychological experience of interacting with human avatars or agents in immersive virtual environments, and (ii) methods and affordances for computer-mediated communication and collaboration. Studies of the psychological experience of interacting with *embodied agents* or *avatars* in immersive virtual environments have focused on *agency, presence, copresence* (or *social presence*), and *social influence* [BY05, Bla02, FAJ+15]. They employ self-reports, behavioral metrics, cognitive metrics, and qualitative methods to gain insight. Two factors shown to influence all of the above are *behavioral realism* and *photorealism* of the agent or avatar representations [BSH+05]. Importantly, a recent meta-analysis [FAJ+15] showed that avatars have greater social influence than agents. That is to say, people react more strongly to other people than to non-human agents that purport to be people. In the present work, we are only concerned with the case of real-time interaction between people, so the upshot is that our use case resides at the end of the spectrum where social influence tends to be larger. A relevant study by Garau et al. [GSV+03] considers this case, also through the lens of behavioral and photo-realism. In their system, users' headsets and a single handheld controller are spatially tracked with six degrees of freedom. The authors define a metric for the *perceived quality of communication*, and test how this depends on *type of avatar* and *type of gaze*. The former refers to three different levels of realism, and the latter refers to two different methods for generating avatar eye gaze behavior. Results show a positive effect when gaze behavior mimics natural behavior. However, the said "natural behavior" is inferred from a model of speaker turn-taking, and not directly controlled by the user's real eye gaze. In contrast, our system does not use any indirect inference: it displays only the head orientation, and does not purport to represent eye movement.

151

It also displays hand positions, supporting the use of unintentional and symbolic gestures.

The related work in the field of computer-mediated communication investigates the merits of different communication affordances from the perspective of collaboration. Isaacs and Tang [IT94] perform a systematic comparison of audio, video, and face-to-face as media for communication. They note increases in communication efficiency in video over audio-only communication due to the ability to indicate agreement using a nodding gesture, without interrupting the speaker. They note the great value of being able to point in the shared environment, as in face-to-face communication, but also highlight that video can be more efficient than face-to-face in cases where it removes distractions. Our system supports nodding to express agreement, and we also make observations about the removal of distractions in our somewhat different setup. In [OMIM94] from the same year, the authors focus on gaze and the representation of video avatars. They contend that the ability to judge which other participant is being gazed upon by each participant is important for group dynamics. Our system also allows each participant to see where other participants are looking through their head orientation, which we confirm to be an important feature. More recently, [KLSB14] uses see-through display augmented reality for remote collaboration. This work considers puzzle-solving as a collaborative task, and also underscores the importance of the affordance for pointing when collaborating in a shared space. Our system supports the ability to point in space, in a way that is directly analogous to the physical world except for the small physical disparity between the user's physical and virtual hands. The most similar prior work from the field of computer-mediated communication is *GreenSpace II* [DC96], a multi-user 6DoF system for architectural design review. Its two users would see stylized head and hand avatars (with one hand per user), and point in the shared space. Their physical movements were constrained to a small space – to make larger movements, they needed to use a 6DoF mouse. The paper demonstrates the feasibility of sharing an immersive virtual environment with spatially tracked head and hand avatars. A significant portion of the feedback provided in the qualitative evaluation focused on the limitations of the technology. The present work does confirm what is supposed there – namely that once the fidelity of the experience is improved (wider field of view, natural physical movement, better audio experience), the utility improves greatly, and the interaction feels natural.

5.2.2 System for Copresence in Room-Scale VR

We present a system to act as a foundation for exploring same-time, same-place collaboration in room-scale virtual reality. A later version of the system, described in Greenwald, et al. [GCM17], is available for the community to use.[1] It allows each user to see head and hand avatars representing the other user, with their apparent virtual positions matching their respective physical positions, as illustrated in Figure 5-12. The form of the head avatar corresponds closely to the physical headset. The hand avatars are customized according to the activity being performed.

The choice not to display a head or a body was made in order to be deliberately minimal – representing the hardware itself, so as to avoid making arbitrary choices that could significantly influence the experience. An entire field of related work (see e.g. [SSV14]) concerns itself with how the representation of the body impacts the user's psychological experience, and we are just concerned with the baseline communication capabilities in the scope of this paper. Even so, we did opt for a few minor tweaks based on the results of preliminary testing. The headset is modified with the addition of simple, static eyes on the front, since users found that this dramatically increased the sense of social presence. In pilot testing, users had difficulty creating expressive hand gestures using a literal representation of a hand holding a controller. Instead, the default hand avatars are flat hands positioned vertically above the top of the controller, which proved to be more versatile.

Our system uses the the HTC Vive, an off-the-shelf 6DoF VR system consisting of a headset, a pair of handheld controllers, and pair of tracking base stations. The Vive system requires one computer per headset, but several systems can share a set of base stations. Sharing is possible because the devices being tracked (headset and controllers) are receivers which observe optical signals from passive base stations. We calibrate a single coordinate system between the VR systems by sharing a set of configuration files between their host computers. Players' apparent virtual locations are made to match their physical locations, and the systems continually synchronize a virtual world representation over a local network. Our "naive" implementation sends updated headset and handheld controller positions from every user to every other user at 90Hz, and has been tested with a maximum of five users in a single space. With that number of users two challenges

---

[1]CocoVerse, https://github.com/cocoverse

arise: (i) with our "naive" implementation, network and graphics performance start to suffer, and (ii) physical cable management, with a cable running to each user's headset. Our environment was implemented in Unity, and we used a custom serialization protocol and TCP connection in the provided networking framework to synchronize the state of the environment between the host computers.

In order to be able to comprehensively study user interactions that take place in our system, we considered it an essential design requirement to be able to record and playback these interactions. Rather than screen recording, which is limited to one or two perspectives, we opted for recording of 3D paths of motion and orientation. This format supports visual inspection and quantitative analysis alike, allowing recordings to be viewed from any angle, and analyzed numerically. Viewing replays of VR interactions while actually in the VR space is a novel and insightful experience, and this topic will be discussed further in our experimental results.

5.2.3 Experiment Comparing Face-to-Face with VR

We sought reference activities to help us accomplish the stated goals of investigating advantages and disadvantages of VR vs face-to-face for communicative interaction, and exploring the use of freehand drawing in 3D in this setting. We identified the word-guessing games *Charades* and *Pictionary* that fit these constraints. These require the use of gestural communication that is both symbolic and expressive, and they are also composed of a sequence of short, goal-oriented sub-tasks exercising different means of non-verbal and gestural communication. Pictionary also has the property of being naturally extensible from its familiar 2D form into a 3D form — allowing for 2D and 3D interactions to be compared side-by-side as well, providing a baseline for investigating free-hand drawing in 3D. In the Charades game, the focus of communication is on the body itself, while Pictionary makes use of a spatial medium to contain and convey drawings. This contrast should yield greater insight into the effectiveness of these two different communicative affordances, body movement and drawing, and allow us to conjecture what kinds of activities would be most amenable to this form of collaboration. It should also help identify the most limiting technological shortcomings, and hence provide recommendations about what improvements would be most worthy of effort. Overall, we see the communicative gestures and actions required by these two different word guessing games as a proxy for the many kinds of communication required for a variety of
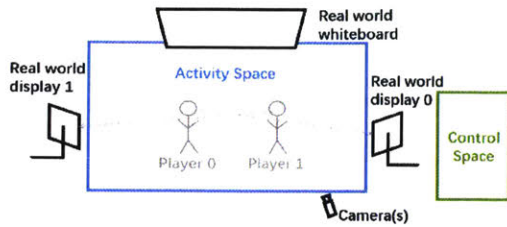
Figure 5-14: Physical room layout for face-to-face and VR games.
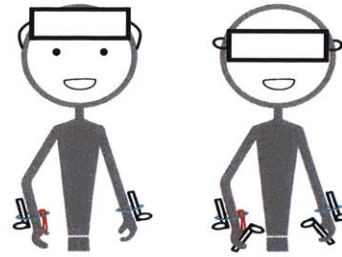
Figure 5-15: Positioning of headset, controllers, and sensors during F2F and VR activities.

collaborative tasks. The tasks themselves are communicative, but only "collaborative" to a limited extent, since only one participant acts at a time. Isolating one-way communication in this fashion will act as a first step, paving the way for future research into more complex collaborative tasks using this configuration.

Method

To compare the effect of these independent variables (face-to-face versus VR conditions, and the two game-based task settings), we conducted a user study. We designed our experiment following a repeated measures design with one independent variable: the word guessing game that is played (Charades or Pictionary) combined with whether the game was played in VR or F2F. As dependent variables we measured the *Electrodermal Activity (EDA)* through sensors, *Task Load Index (TLX)*, *level of presence* as well as some other related aspects of the system usability through questionnaires. We counter-balanced the order of the conditions according to the Balanced Latin Square.

The two primary hypotheses related to the contrast between our independent variables were that (1) face-to-face and VR would be similarly effective, despite the ostensible differences in the richness of the communication channels, and (2) the games would reveal different quantitative and qualitative attributes of non-verbal communication across conditions, given their different uses of body movement versus drawing.

Apparatus and Tasks

Room setup. Figure 5-14 depicts the physical space layout used for the experiment. Players act or draw in the *activity space*. Facilitators operate and control the game session from the *control*

155

space. The *real-world whiteboard* is used for drawing during the F2F Pictionary game. Game play information such as the current word, timer, game mode etc. is shown on the *real-world display* during F2F conditions. The *camera* footage of the F2F games provides video for think-aloud review sessions.

Positioning of devices on body. Figure 5-15 shows the positioning of the GSR sensors, VR controllers and headset on the body during F2F and VR activities, mounted with elastic velcro bands. The positional tracking devices worn during the activities collected movement data that could be directly compared between F2F and VR conditions. The GSR sensor was mounted to participants' dominant hand, with gel electrodes placed on the lower palm.

Quantitative data acquisition. Electrodermal activity data was collected using a *Shimmer GSR* sensor with *iMotions* software. After smoothing and detrending, *Coefficient of Variation (CV)* was calculated as a metric of arousal, following Doberenz et al. [DRW+11].

Movement data was collected from the position of the headset and two arm-mounted controllers. The HTC Vive system provides positional data at a rate of 90Hz. The sensors occasionally become momentarily occluded, causing tracking to be lost. We computed the average distance traveled per tracked frame (cm/frame) for each session and player.

Word selection for guessing games. The guessing words used during the study were selected from lists of varying difficulty provided by a game website.[2] For each game, we informally piloted candidate words, and observed the type of body gestures used while playing (fingers, hands, full-body, etc.), as well as the use of 3D space where applicable. Based on the results, we selected a final set of words of varying difficulty that would sample a variety of gesture types and highlight different uses of 3D space.

Questionnaire design. We used the NASA Task Load Index (TLX) questionnaire and a custom set of questions. The TLX questions were presented using a slider with options from 0 to 100 in increments of 5, with the slider initially positioned at 50. Informed by our pilot tests, additional questions were presented to inquire about specific aspects of game play, the differences between F2F and VR, and the usability of the user interface.

---

[2]The Game Gal, https://www.thegamegal.com/

Procedure

Subjects arrived in pairs, and experimental sessions began with a general introduction, before putting on VR devices and sensors. Pairs played through all five conditions (Charades and Pictionary 2D, each in F2F and VR, plus Pictionary 3D in VR) in the order dictated by their experimental group, with questionnaires administered as appropriate after each condition. At the start of each condition, participants were first given an opportunity to briefly familiarize themselves with the devices and physical or virtual space, and a simple warm-up task was provided. During game play, for each word the "acting" player was given 45 seconds to silently convey a word to the "guessing" player, with roles alternating as directed by the system. The facilitator determined when the word had been guessed correctly, and operated a control interface on one host computer to advance to the next word. After playing both F2F and VR variants of a game, participants would perform a retrospective think-aloud protocol and interview together. They reviewed the video and immersive VR playback (or just immersive VR playback, in the case of Pictionary 3D) in succession, in the order that they were played.

Participants

We invited 6 pairs of participants (4 female and 8 male) to take part in the study, with ages ranging from 19 to 50 ($M = 31.0$ years, $SD = 10.62$ y). The study took approximately 2.5 hours, of which roughly 30 minutes were spent playing the games, 30 minutes reviewing recordings, 30 minutes filling out questionnaires, 30 minutes interviewing, and the remaining time used for breaks and setup. Participants were compensated with a $25 gift card.

Quantitative Results

Here we present the data that was collected during the user study. To analyze the NASA-Task Load Index (TLX), we used a one-way repeated measures ANOVA. For the questionnaire we applied a non-parametric Friedman test. Bonferroni correction was used for all post-hoc tests.

NASA-TLX When comparing the TLX between the five conditions, the F2F Charades led to the least perceived cognitive load ($M = 49.33$, $SD = 18.6$), followed by the VR Charades ($M = 50.17$, $SD = 10.26$), the 2D VR Pictionary ($M = 60.34$, $SD = 9.55$), the 2D F2F Pictionary ($M = 60.58$,

$SD = 9.26$), and the 3D Pictionary ($M = 66.17$, $SD = 9.60$). Mauchly's test of sphericity indicated that we can assume a sphericity of the data ($p > 0.05$). The one-way repeated measures ANOVA revealed a significant difference between the conditions, $F(1, 4) = 6.589$, $p < .001$. As a post-hoc test, pairwise comparisons revealed a significant difference between the VR Charades condition and the 3D VR Pictionary condition ($p < 0.05$). The effect size shows a large effect ($\eta^2 = .375$). Figure 5-16a shows the results graphically.

(a)   (b)
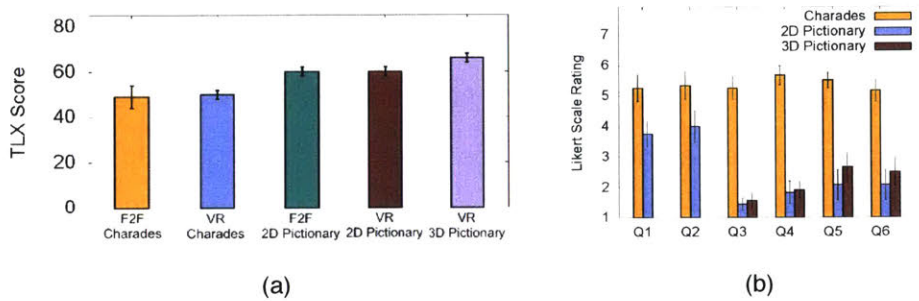
Figure 5-16: (a) The NASA-Task Load Index results of the user study for all conditions and (b) The quantitative results of the Likert scale questionnaire for the different games. Questions Q1-Q6 are explained in the text. All error bars depict the Standard Error.

When analyzing the Likert questions of the questionnaire, we used a non-parametric Friedman test. All Likert items were 7-point Likert items meaning: 1 = strongly disagree and 7 = strongly agree. For Q3-Q6 we used Wilcoxon signed-rank post-hoc tests with an applied Bonferroni correction for all conditions resulting in a significance level of $p < 0.017$. All results of the questionnaire are depicted in Figure 5-16b.

Q1: *"Overall the experience playing the game in VR was different than playing F2F."* Participants found that the overall experience playing the Charades game in VR was more different from playing it F2F ($M = 5.25$, $SD = 1.49$) than in the 2D Pictionary game ($M = 3.75$, $SD = 1.42$). The Friedman test revealed a significant difference between the two games, $\chi^2(1) = 6.0$, $p = 0.014$.

Q2: *"Playing the game in VR was harder than playing F2F."* Further, the participants rated playing the Charades game to be harder in VR compared to F2F ($M = 5.33$, $SD = 1.56$), compared to the 2D Pictionary game ($M = 4.00$, $SD = 1.76$). The Friedman test did not reveal a significant difference between the two games ($p > 0.05$).

Q3:*"The absence of a body avatar was a problem in VR."* Considering the absence of a body avatar, the participants the participants rated the Charades game the most problematic (M = 5.27, SD = 1.35), followed by the 3D VR Pictionary (M = 1.55, SD = .93), and the 2D VR Pictionary (M = 1.45, SD = .69). The Friedman test revealed a significant difference between the games, $\chi^2(2)$ = 18.0, $p < 0.001$. The post-hoc tests showed a significant difference between 2D Pictionary and Charades (Z = −2.825, p = 0.005) and 3D Pictionary and Charades (Z = −3.072, p = 0.002).

Q4:*"The absence of facial gesture representations was a problem in VR."* When analyzing if the absence of facial gesture representations were a problem in VR, the participants rated the Charades game as the most problematic for that aspect (M = 5.67, SD = 1.07), followed by the 3D VR Pictionary (M = 1.92, SD = .9), and the 2D VR Pictionary (M = 1.83, SD = 1.27). The Friedman test revealed a significant difference between the games, $\chi^2(2)$ = 20.14, $p < 0.001$. The post-hoc tests showed a significant difference between 2D Pictionary and Charades (Z = −3.075, p = 0.002) and 3D Pictionary and Charades (Z = −3.089, p = 0.002).

Q5:*"The absence of hand gesture representations was a problem in VR."* Considering if the absence of hand gesture representations is problematic in the VR games, the participants rated the Charades game as the most problematic (M = 5.5, SD = .905), followed by the 3D VR Pictionary (M = 2.67, SD = 1.67), and the 2D VR Pictionary (M = 2.08, SD = 1.73). The Friedman test revealed a significant difference between the games, $\chi^2(2)$ = 17.077, $p < 0.001$. The post-hoc tests showed a significant difference between 2D Pictionary and Charades (Z = −2.842, p = 0.004) and 3D Pictionary and Charades (Z = −2.952, p = 0.003).

Q6:*"The absence of finger gesture representations was a problem in VR."* Finally, when analyzing whether the absence of finger gesture representation was problematic for playing the VR game, the participants rated the Charades game as the most problematic (M = 5.17, SD = 1.267), followed by the 3D VR Pictionary (M = 2.50, SD = 1.567), and the 2D VR Pictionary (M = 2.08, SD = 1.73). The Friedman test revealed a significant difference between the games, $\chi^2(2)$ = 11.73, p = 0.003. The post-hoc tests showed a significant difference between 2D Pictionary and Charades (Z = −2.739, p = 0.006) and 3D Pictionary and Charades (Z = −2.823, p = 0.005).

Considering the players' analysis of their experience in both games we were asking additional

questions comparing their VR and F2F experience.

Q7: "Reviewing videos/the VR recordings helped me remember my experience during the games." When analyzing where the participants found it better to review their experience, the participants found the VR recording of the games better ($M = 6.00$, $SD = 1.27$) than the video recording ($M = 5.58$, $SD = .51$). A non-parametric Friedman test could not find a significant difference between the video recording and the VR recording.

Q8: "Reviewing videos in VR/on video helped me gain new insights into my interactions". Considering gaining new insights on the participants interactions during the game, the participants rated the VR recording to provide more insights ($M = 6.00$, $SD = 1.20$) compared to the traditional video recording ($M = 4.91$, $SD = 1.37$). A Friedman test revealed a significant difference between the two recording systems, $\chi^2(1) = 4.500$, $p = 0.034$.

Electrodermal Activity Considering the analysis of the EDA, the results revealed that the 3D VR Pictionary led to the most EDA activity ($M = .24$, $SD = .18$), followed by the 2D VR Pictionary ($M = .20$, $SD = .20$), the F2F Charades ($M = .15$, $SD = .11$), the VR Charades ($M = .14$, $SD = .07$), and the 2D F2F Pictionary ($M = .11$, $SD = .04$). Mauchly's test of sphericity indicated that we cannot assume a sphericity of the data ($p < 0.001$). Therefore, we apply a Greenhouse-Geisser correction to adjust the degrees of freedom. Unfortunately, a one-way repeated measures ANOVA could not reveal a significant difference between the conditions ($p > .05$).

Head Movement When analyzing the head movements the participants made during the different conditions, the 3D Pictionary ($M = .117$, $SD = .038$), Charades F2F ($M = .115$, $SD = .045$), and the Charades VR ($M = .111$, $SD = .041$) lead to similarly frequent head movements, followed by the F2F Pictionary 2D ($M = .105$, $SD = .045$). The Pictionary 2D in VR led to the least head movements ($M = .078$, $SD = .018$). A one-way repeated measures ANOVA revealed a significant difference between the conditions, $F(4, 32) = 2.670$, $p = .049$. However, a post-hoc did not reveal a significant difference.

Left Hand Movement For the movements of the participants' left hands, we found that the F2F Charades led to the most hand movement ($M = .27$, $SD = .15$), followed by the 2D F2F Pictionary ($M = .21$, $SD = .10$), the VR Charades ($M = .19$, $SD = .05$), the 3D Pictionary ($M = .13$, $SD = .04$), and the 2D VR Pictionary ($M = .10$, $SD = .03$). Mauchly's test of sphericity indicated that we cannot assume a sphericity of the data ($p < 0.001$). Therefore, we apply a Greenhouse-Geisser

160

correction to adjust the degrees of freedom. A one-way repeated measures ANOVA showed a significant difference between the conditions, $F(1.477, 14.771) = 8.775$, $p = .005$. The post hoc tests showed a significant difference between 2D VR Pictionary and all other conditions. Further there was a significant difference between VR Charades and 3D Pictionary (all $p < .05$).

Right Hand Movement   We found that the 3D Pictionary led to the most right hand movement ($M = .26$, $SD = .12$), followed by F2F Charades ($M = .24$, $SD = .11$), the 2D VR Pictionary ($M = .23$, $SD = .18$), the 2D F2F Pictionary ($M = .23$, $SD = .11$), and the VR Charades ($M = .21$, $SD = .06$). A one-way repeated measures ANOVA could not reveal a significant difference between the conditions ($p > .05$).

Qualitative Results

The questionnaire questions reported above captured many of the most salient trends we discovered during our prior informal pilots. The qualitative results presented in this subsubsection are focused on ideas that are either more complex and nuanced, or first became apparent in the main study. In this section we report factual aspects of this feedback, and save a discussion of its significance and relationship to our quantitative results for the *Discussion* section that follows.

One idea that was important but also very subtle to interpret was the degree of expressivity participants perceived in the gestures of others. This subject was always brought up in the interview at the end of the entire session. All participants agreed that, as expected, the smoothness and precision of the representation of movement in the space led to a high degree of expressivity and sense of being able to perceive some aspects of emotion or other non-verbal reactions. It was difficult for participants to describe this explicitly, because in the same-time, same-place setting, it seemed very natural that the other person's emotions could be interpreted through movement, and therefore not noteworthy on its own. For this reason it was primarily during the process of viewing VR recordings that participants were able to consider in isolation what kind of information avatar movements contained. Several participants found their own movements and those of their partners to be distinctive and recognizable. Other participants disagreed, and felt that they would not be able to distinguish a playback of their own avatar actions from actions of unknown others. This on-the-fence status was well summarized by one participant's comment that there were "glimpses of humanity" that would appear sporadically throughout the process of viewing. Another participant

reported "they're very emotive" and "you can definitely tell it's you."

Recounting briefly some comments about the general relationship between the face-to-face and VR experiences, participants mentioned most frequently that VR Charades was challenging because of the lack of face and body avatars. After initial reports that the VR 2D Pictionary experience was qualitatively highly similar to its face-to-face counterpart, the facilitators questioned participants for more detail. Because participants rarely look to each others' faces for feedback during gameplay, the entire focus was really on the board, and they found the experience of drawing on the physical whiteboard versus the virtual whiteboard nearly identical. They cited several advantages for VR over face-to-face: the virtual board erases automatically between words, switching colors was faster using the VR color palette than physically switching markers, and in VR the body does not occlude the drawing surface, so it was never an issue that the actor's body was blocking the view. One corollary that came out in interviews was that VR offered the advantage of removing some aspects of face-to-face interaction that are distracting, awkward, or unpleasant. Attention to gender, ethnicity, body image, and certain visual social cues are impeded through the invisibility of the physical body.

Next, we review comments participants made about the process of reviewing video versus VR recordings. Several participants reported reviewing video to be unpleasant, mentioning they felt "silly" watching themselves play. In contrast, they described the experience of watching replays in VR as insightful and fun. In 3D Pictionary specifically, many participants reported that viewing the replay from a different perspective allowed them to see how their drawings were not as decipherable from their partners' perspective as from their own.

One last area of participant feedback that we'll highlight in this section is the description of 3D versus 2D drawing. Nearly all participants described drawing in 3D as challenging, but some enjoyed the challenge while others found it frustrating. There was broad agreement that drawing in 3D was typically slower, but there were cases where it offered advantages. The biggest challenge was becoming accustomed to considering multiple viewing perspectives. There was a weak consensus that drawing on a virtual 2D plane would be a winning strategy if emphasis was placed on finishing quickly. In contrast, participants in our experiment participants were given time limits, but were not

otherwise incentivized to finish quickly. This observation is highly coupled to the specific task of Pictionary play, and may have been accentuated by the fact that the word list was designed for 2D Pictionary.

Figure 5-17: Expressive poses in VR/F2F acting out "blind" (left) and "beg" (right)

5.2.4   Discussion

The previous section presents a disparate set of results from our five data sources. In this section we highlight some salient relationships between these results.

We begin by observing that participants felt strongly that (1) the communication medium was not sufficient for Charades, while feeling that (2) the medium was entirely sufficient for Pictionary in 2D and 3D, as evidenced by the questionnaire responses. In the former, the absence of facial gestures, finer hand gestures, finger movements, and a body for non-verbal communication were considered highly problematic, while in Pictionary they were considered irrelevant. Further underscoring this was the response to Q1. At the Likert scale value of 3.75 participants were very close to "neutral" on the question. We interpret this as a strong statement about two aspects of the interaction: (1) the adequacy of the hand-held controllers at approximating the face-to-face experience of drawing on a whiteboard, and (2) the expressiveness of the avatars. We know that when the focus of the interaction is on the body itself, as in Charades, the simple avatars were inadequate. Despite participants' reports to this effect, even the most difficult words we tested were guessed correctly by a subset of groups – meaning that the communicative affordances were nonetheless powerful enough to admit creative workarounds. Furthermore, the qualitative feedback indicated that the avatars were perceived as quite expressive and emotive. Reconciling these statements, we propose the following guideline, pertaining to systems equivalent to ours: a collaborative task that is communicative, but with a central focus that is not on the face or body itself, when facilitated by well-adapted task-specific interface affordances, will yield an overall experience comparable

163

to face-to-face. Stated more broadly, minimal avatars provide a powerful and versatile baseline set of communication affordances. Roughly speaking, the two games we tested define a spectrum between the worst and best-adapted activities for our simple head and hand avatars. We conclude that, when designing system for a certain form of collaboration in VR, one should ask whether it is more Charades-like or more Pictionary-like in order to decide whether the additional effort of embodying a more sophisticated avatar is justified.

Next, comparing movement, TLX, and EDA data for 2D Pictionary reveals an interesting correlation. In particular, it was a high-EDA activity, and a somewhat high perceived cognitive load (TLX) activity, while being the lowest-movement activity overall. This indicates a mode of mental engagement corresponding to decreased physical movement. If there were any coupling between physical movement and EDA, it would work against this result, hence it is interesting to highlight.

Now we review true advantages of VR over face-to-face that were shown in our results, beginning with those relating to efficiency of task performance. First, the virtual whiteboard did not need to be manually erased, and therefore decreased the time and energy required to perform an equivalent task in VR versus F2F. Next, the transparency of the body in VR minimized occlusion of the virtual whiteboard – the drawing player could stand right in front of the board without preventing the guessing player from seeing the drawing. Next, a psychological benefit was reported in participants' observation that masking the physical body can be beneficial to focus and decrease social anxiety in collaborative interactions. All of these can be viewed as advantages of "programming" the virtual visual environment, by instantly changing its properties in ways that require time and effort, or aren't possible at all, in the physical world. Indeed, they satisfy physical and psychological needs of communication in a way that is not possible face-to-face, and hence go *beyond being there* [HS92].

Now we turn briefly to the methodological implications of this experiment. Although our EDA data did not uncover significant differences between our activities, it was close enough that we would conjecture that further refinement of the method to reveal significant differences would be possible – for instance subdividing overall games into smaller components, or applying peak detection algorithms. Next, discussing movement data, the only significant result was that the left (palette)

164

hand stays very still during 2D Pictionary. While this is not exciting on its own, the prospect of doing more sophisticated analysis of body movement with absolute positional data rather than (or in addition to) accelerometry seems promising. This is firm evidence that activity analysis and recognition can be applied to the positional data collected by the Lighthouse system, and certainly any other system with similar or greater precision that comes along. Finally, our significant result about the difference between video and VR review of games is worthy of note. Participants found VR review equally good (i.e. not significantly different) for *recall* of the experiment, but significantly better at providing new insights. Not only does this provide a basis for researchers to obtain highly nuanced qualtitative feedback from participants, it also suggests that review of VR activities could be used in the context of learning or training – leveraging the reflective power of scrutinizing ones' own performance in a way that is demonstrably better than video.

5.2.5 Conclusion

Same-time, same-place interaction in virtual reality has been shown without any doubt as a practical medium for communication and collaboration, which carries with it a sense of social presence that is adequate for a variety of non-verbal methods of communication mediated by hand gestures, head gestures, and overall spatial movement. This corroborates the observations I reported in Section 4.2, in which people became highly engaged with the communicative and collaborative aspects of their interactions in CocoVerse. Adding a finer point to the statement, if facial gestures, torso, or leg movements are particularly relevant to the communicative task, the minimal system we built would need to be extended to support these in some fashion before being applied for the use case. It was shown that drawing in 3D is challenging but highly promising due to the new space for expression that it opens up, that has no physical analog. It was observed that interacting in VR has the advantage of masking aspects of physical appearance and the body that can be distracting during collaborative interaction. Reviewing interaction in VR allowed participants to gain new insight into how their own communicative processes did and didn't work, and this could be useful as a tool for reflection or coaching. I see all three of these as fruitful directions for future research in collocated and remote computer-mediated communication using room-scale virtual reality.

5.3 Summary of Experimental Studies

In this chapter, I presented two experimental studies: the first related to the personal or environmental purposes of the Equipped Explorer (Section 5.1), and the second to the social purposes of communication and collaboration. In both cases there was a hypothesis that VR has a special ability to create meaningful and impactful experiences: in the case of the VR versus 2D learning study, it was that interactive an VR experience would lead to better comprehension than a 2D alternative; and in the case of the social presence study it was that people could sense emotive human social signals and non-verbal communication through minimal avatars in 6DoF VR. The latter derives its impact not by way of facilitating communication on its own– which can be done without a VR headset– but through what it enables people to experience together, combining a meaningful human experience with novel interactions (like painting in 3D) and environments. In both cases these basic hypotheses were confirmed, and perhaps even more importantly, a great deal of insight was gained into the properties of the respective design spaces, and methods for studying such interactions. In the VR learning study, VR was shown to lead to greater comprehension than a 2D alternative, although this result was obtained for only one of two activities. Some of the deeper discoveries concerned design guidelines for VR learning activities, and ways of using behavioral metrics to determine what kind of learning is taking place. In the social presence study, participants showed they could communicate a variety of complex ideas just using gestures in VR, and uncovered challenges and opportunities: 3D drawing is challenging but also highly expressive, and having a minimal avatar can remove distractions from collaborative interactions. These results will be considered in the context of the entire dissertation in the the next and last chapter.

Chapter 6

Conclusion and Outlook

The four key contributions of this dissertation are (1) a review of background literature, which lays out concepts in cognitive science that point to advantages to learning in VR (2) a two-part design framework, with the Equipped Explorer as a metaphor and framework for classifying and designing for exploratory learning activities, and Reality Integration as a taxonomy of needs for understanding how such experiences can be "plugged in" to real-world contexts; (3) a series of design explorations that investigate the learner's experience vis-à-vis the virtual environment, the learner's experience in the virtual environment vis-à-vis other people sharing the environment, and ways of implementing certain key Reality Integrations; (4) a pair of experimental studies, investigating the learning benefits of activities in VR compared with 2D, and the ability to have meaningful human experiences when interacting with others in VR. In the course of the four corresponding chapters, countless interesting facets of learning in VR emerged, but until this point in the dissertation, these have not yet been related to one another to create an overall picture. That is one goal I hope to accomplish in this final chapter. I will also present a few new ideas expounding on some of the most interesting findings, and conclude with thoughts on the broader implications and most important directions for future work.

6.1 Recap

In the Background chapter, one observation that frames the discussion of the advantages of VR is that VR must be considered distinct from a 3D display. The primary distinction is that the motion

of the head controls perspective, and the hands allow the user to directly interact and manipulate elements in space. With that as a basis, I would first highlight the claim that VR allows multiple representations to be combined in one place. The combination of stereoscopic presentation with the intuitive and instantaneous changing of perspectives afforded by head movement allows the combination of representations to take place without being visually cluttered or confusing. Drawing electric field lines in 3D space as in the Electrostatic Playground is an excellent example of this. The TerrImmerse application is another example where the combination of hand-based interaction to move and scale the content, with the ability to move the head allows the user to more easily make sense of 3D phenomena like intersecting planes, and two different overlaid representations that can be displayed in rapid alternation. The second claim I would like to highlight is that VR affords verbatim note-taking without incurring cognitive load (as would audio/video recording), but suffers none of the loss of quality that audio and video recordings do. I refer in particular to the ability to see new perspectives when reviewing learning experiences, which avoids losing critical information due to occlusion or field of view. Bookmarking helps to make such recordings more efficiently usable by making moments quickly accessible where the learner has either good insights or lingering questions. The Neuron and Safari applications pilot such interfaces, specifically in the context of spatial exploration of 3D models related to biology.

In the Design Framework chapter, the Equipped Explorer builds on the concept of exploratory learning introduced in the Background chapter. Thinking about learning through the lens of interaction tools puts the learner in control of the experience, and hence is compatible with the constructivist theory of learning. The four categories I propose: exploration, communication/collaboration, capture, and review are strongly informed by Kolb's cycle [Kol84] which proposes that learning takes place in a sequence of events of experience, reflection, abstraction, and active experimentation. The use of these categories allows the prototypes presented in the following chapter to achieve coverage of the most important purposes an exploratory learner might have. The Reality Integration taxonomy gives structure to the set of considerations that arise when embedding the use of VR in real-world situations– from safety, to creating pipelines for data to come in and out of the VR experience, to allowing people outside VR to act as facilitators for the experience of people inside VR. The broad categories are spatial, temporal, and contextual integration. Over time, the

hope is that providing this taxonomy will give people a way of thinking about the coverage of these needs. Designers of commercial hardware and software products, for example, can use these to identify the different needs that their products can and must support to maximize the utility and impact of the technology. The multiple iterations on the CocoVerse environment provide examples of temporal, spatial, and contextual Reality Integration requirements that are solved through software implementation. The Window project shows how the use of an additional hardware device can add the versatility to support many practical use cases spanning the space inside and outside VR.

Next, in the Prototypes chapter, I discussed a variety of applications based around certain content, but always with the goal of abstracting beyond the content: reasoning about the reusability of tools, and striving to address in the most general possible way the design challenges that arise. In the section on interactions between the learner and the environment, some of the thorniest and most interesting challenges were related to designing interactions around simulations. My concept of exploratory learning was conceived in thinking about interacting with simulations– although its applicability is not limited to them– so it is not surprising that many of the prototypes involved them. Two categories of challenge that arose were (1) how to design for interactivity, especially high-bandwidth spatial interactivity, when the target subject matter seems to resist it, and (2) how to deal with the fact that there are boundaries to the realism of every simulation.

Next, when considering collocated multi-user (social) applications, the CocoVerse application proved unexpectedly captivating, and a substantial amount of analysis in Section 4.2.3 was devoted to documenting their behaviors and understanding why. In particular, people became highly engaged, and would easily use the application until some outside constraint stopped them– anywhere from 20 to 60 minutes. Understanding what made this world so engaging, I hope, will allow others to design such environments with learning goals cleverly woven in. One high-level takeaway, however, is that the combination of meaningful human communication with virtual worlds appears to be very powerful.

Moving on to the Experimental Studies chapter, the goal is to put the raison d'être of the prototypes on more solid footing, by showing that learning in VR has cognitive advantages, and that the social

scenarios I propose succeed at providing a high level of human communication and connection. The study on VR physics learning activities revealed that learning in VR could change how learners approached a 2D activity, and ultimately lead to better comprehension of the subject matter. This advantage was observed in only one of two activities, and my hypothesis is that the design of activities in VR must be well-adapted to VR. Unlike in the first activity, learners found the 3D aspect of the second activity confusing and didn't find the insights revealed by it to be intrinsically interesting or motivating. Similar challenges exist in designing 2D activities as well, and the shortcomings of our second activity appear to be foreseeable given a set of design guidelines. The goal now is to quickly turn this from a burgeoning craft shaped by tacit knowledge to a design science with clear guidelines, and the results from this study provide a starting point for such guidelines.

The study on social presence in VR yielded the general insight that the communication afforded by minimal avatars fulfills most purposes of communication for learning and collaboration where there is not a specific focus or need to represent the parts of the body that are missing from the avatar (e.g. fingers, mouth, feet). It has the unexpected advantage of removing distractions related to physical appearance, as well as reducing problems associated with occlusion of shared resources in a face-to-face setting. One interesting and important side-finding from this study concerned the properties of the retrospective think-aloud protocol we implemented. Because people watched their prior activities *in VR*, they had the opportunity to view the scenario from different perspectives. This provided a variety of new insights: participants could see what their partner was seeing, which allowed them to better understand their behavior, and also allowed them to reflect on their own strategy and behavior. In the next section, I will further consider the implications of this finding.

6.2 The Power of VR Recording

In different ways, numerous prototypes and both experimental studies presented attested the power of VR recording. It promotes reflection and gaining new perspectives on learning experiences, enables an entire category of constructionist learning activities, and empowers teachers to create narrative and interactive content without the need to program. In this section, I will reflect on these findings and their ramifications.

6.2.1 Reflection and Metacognition

As highlighted above, participants' process of watching a replay of events in VR revealed something unexpected about the medium itself. I had mentioned in Section 2.2.5, that recording in VR should allow for the generation of high-quality verbatim "notes" in the form of an immersive capture of a learning experience. I also presented two prototypes that were based around the idea of intentional capture of positions in space and orientations of objects (Neuron and Safari in Section 4.1.3), and a tool for recording in Section 4.3.2. However, the unanticipated characteristics of verbatim notes revealed themselves in the experimental study on social presence (Section 5.2), when the content being reviewed was dynamic and contained the participants' own avatar along with those of others they were interacting with. This behavior was seen as very expressive, and even more importantly, it allowed the participant to gain insight into their own behavior and the perspective of others on subsequent viewings.

6.2.2 An Ecosystem of Content Based on VR Recording

Combining the active and passive uses of recording described above should give rise to a rich, multi-faceted ecosystem of content. Producing content can be a constructivist learning activity on its own; materials that reveal how others' misconceptions are corrected can be captured passively; and teachers can create immersive and engaging learning experiences without programming by recording explanations or setting up interactive scenarios targeted at particular learning goals. Others can easily benefit from the byproducts of these processes, since such content is as easy to share as a any electronic document. As part of playful learning experiences, learners might, for example browse for the best, funniest, or most clever examples of other students doing similar activities.

Some of the attributes of VR recordings just mentioned– purportedly critical enablers for a new kind of learning content ecosystem– may at first glance appear to also apply to existing media, like webpages and 2D video. However, there are important differences that provide much greater friction for these alternatives. First, let us consider the authorship and sharing characteristics of webpages. Making webpages requires rudimentary familiarity with programming in all but very special cases. Even in cases where it is easy to make (e.g. through the use of a WSIWYG content

171

management system), it is almost never a passive byproduct of teaching, and this creates a barrier to content creation. On the positive side, it seems that it is indeed common for teachers to quickly identify and use web content obtained by searching. That is, the barrier is less on the side of consumption and more on the side of production. Lastly, consuming the content of a webpage is not a particularly engaging experience on its own for most learners.

Now, let us consider 2D video recording. As highlighted previously, viewing 2D video content is sometimes onerous, especially for children, and video recordings are typically only usable when an effort is made to accommodate their vantage point. This disrupts the learning process and precludes the passive creation of useful content. Creating passive VR recordings, in contrast, does not alter the learning experience, and captures all vantage points. To summarize, VR recordings are both more exciting to watch and easier to make than 2D recordings. There is, however, a particular exception to these generalizations about 2D video, that I argue provides evidence that the characteristics I have identified are important ones.

Today, it is very popular for children to watch 2D videos of others playing video games (one of the most popular is Minecraft[1], an open-world sandbox game). According to the above logic, 2D video content of others' activities is too hard to make, and not exciting enough. So why should videos of people playing Minecraft be an exception? There are three special properties I will highlight that explain this phenomenon in terms of the properties I named above: (1) the content is captured passively (which leads to abundance), (2) the camera angle is always well chosen (first-person point of view), and (3) the content itself is engaging, because people enjoy doing the activity being portrayed themselves. It is not coincidental that these characteristics are shared with VR: the first two follow from the fact that the entire experience is mediated by the computer, just like VR.

With the example of Minecraft videos in mind, it does not seem far-fetched to envision that an active global ecosystem of examples, illustrations, explanations, and so on, might be built around VR recording capabilities. Furthermore, an analysis of the cited attributes of existing media attests the validity of the above logic, revealing why, despite the ubiquity of web technology and some of its similar properties, such an ecosystem has not arisen thus far.

---

[1] https://minecraft.net

6.3 Summary and Outlook

I began my dissertation research with a belief and a conviction that VR could accelerate and improve how people learn. I can say at the conclusion that I am as convinced as ever of this. But I did not come all this way just to "state the obvious." From start to finish, I was confronted with the question: is the captivating experience of VR truly beneficial to learning when compared with alternatives? Or might it be only superficially different from what is offered by desktop and mobile computing technologies? There is every indication that VR has the potential to facilitate faster, deeper learning, but as with any other learning technology, it is difficult to quantify its impact. Even so, I did my part to stalwartly quantify it to the best of my ability as part of this dissertation, and I did find a measurable benefit. However, ultimately the paradigm of this dissertation is a design science one, and so the most interesting outcomes are prescriptive in nature: when building a learning application, should one build it in VR? When building a learning application in VR, how should one design it? My findings answer these questions in numerous ways, which I will now summarize.

6.3.1 How should one design VR learning experiences?

Should one build it in VR?

The three best reasons that I have uncovered to build a given learning application in VR would be (1) it involves complex spatial and/or dynamic concepts that are difficult to visualize and comprehend, (2) it involves a significant component of expressivity or creativity, and (3) it involves close collaboration between people.

How should one design it?

I address the question of what attributes the experience should have (as opposed to what design process one should apply). I would conclude here with three design principles:

- **Merge Representations.** Multiple different spatial representations of phenomena can be presented in one place and in situ– no longer should there be a need to glance back and forth between two representations of a system, and piece together in our minds the single

system that the two describe in different ways. That single representation can be accessed directly using VR.

- **Free the Flow of Exploration.** Exploratory learning can be a syncopated process– using 2D input devices with 3D content puts a barrier between imagining and realizing a desired outcome. Namely, a given outcome must be mentally translated from its natural 3D representation into a series of input device actions needed to realize it. This slows down the flow of exploration. VR experiences can free the flow of exploration if they are designed to do so.

- **Leverage the Recordable Environment.** VR gives you the ability to record the entire visual and auditory experience of the learner. Let the learner wield this power to explore, capture, and review.

- **Learn (Optionally) Together.** VR enables people to communicate and collaborate in new ways and in new environments. Let them, but don't force them.

These align with my design contributions related to cognitive advantages of learning in VR; how to design for learning and interactivity when using simulations; how to use recording for recall, reflection, and content creation; and how to create social VR experiences which allow learners to explore independently.

6.3.2 Key Open Questions

The following are some key "obvious" big open questions:

- When is VR advantageous, and how can we focus content development around these areas?

- How can VR be integrated into learning practices, at physical institutions and online?

At this point in time, it appears that these will not be answered all at once, but through the diligent work of many interdisciplinary teams deploying learning experiences in real-world settings. Some equally important questions that were less obvious to me when I began my dissertation are:

- How can the power of recording promote reflection and other metacognitive learning strategies?

174

- How can we enable instructors and students to create content without requiring a software developer?

- What is required to create an ecosystem where this content can be easily exchanged and made useful to others?

6.3.3 Closing Thought

As with any technological innovation applicable to education and the learning sciences, it is prudent to treat VR with a balance of optimism and skepticism. On the optimistic side, we speculate about the advantages of VR, and make good-faith attempts to best leverage its unique abilities. On the skeptical side, we demand and seek evidence that our enthusiasm is grounded; that the technology is really worth the time, effort, and money required to adopt it. Based on the research I've presented here, my optimism has not waned in the slightest. Even so, I remain resolved to maintain the said balance, moving forward with the endeavor of furthering grounded design science related to interactive experiences and the deployment thereof in the area of VR for learning.

Bibliography

[Ain06]   Shaaron Ainsworth. Deft: A conceptual framework for considering learning with multiple representations. *Learning and Instruction*, 16(3):183 – 198, 2006.

[AJRl98]   Edward Awh, John Jonides, and Patricia A. Reuter-lorenz. Rehearsal in spatial working memory. *Journal of Experimental Psychology: Human Perception and Performance*, pages 780–790, 1998.

[BA73]   J.S. Bruner and J.M. Anglin. *Beyond the Information Given: Studies in the Psychology of Knowing*. W W Norton & Company Incorporated, 1973.

[Bad92]   A. Baddeley. Working memory. *Science*, 255(5044):556–559, 1992.

[Bad07]   A. Baddeley. *Working Memory, Thought, and Action*. Oxford Psychology Series. OUP Oxford, 2007.

[BB]   Jim Blascovich and Jeremy Bailenson. Immersive virtual environments and education simulations.

[BF74]   J. A. Birt and T. A. Furness. Visually coupled systems(weapon system integration with operator visual and motor skills). *Air University Review*, 25:28–40, 1974.

[BFO92]   Michael Bajura, Henry Fuchs, and Ryutarou Ohbuchi. Merging virtual objects with the real world: Seeing ultrasound imagery within the patient. In *Proceedings of the 19th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '92, pages 203–210, New York, NY, USA, 1992. ACM.

[BKJJLP01]   Doug A. Bowman, Ernst Kruijff, Jr. Joseph J. LaViola, and Ivan Poupyrev. An introduction to 3-d user interface design. *Presence: Teleoperators and Virtual Environments*, 10(1):96–108, 2001.

[BL91]   S. Bryson and C. Levit. The virtual windtunnel-an environment for the exploration of three-dimensional unsteady flows. In *Visualization, 1991. Visualization '91, Proceedings., IEEE Conference on*, pages 17–24, 407, Oct 1991.

[Bla02]   J Blascovich. Social influence within immersive virtual environments. *The social life of avatars*, 2002.

[BLB+02]   J Blascovich, J Loomis, AC Beall, KR Swinth, and et al. Immersive virtual environment technology as a methodological tool for social psychology. *Psychological*, 2002.

[BOYBJK90] Frederick P. Brooks, Jr., Ming Ouh-Young, James J. Batter, and P. Jerome Kilpatrick. Project gropehaptic displays for scientific visualization. In *Proceedings of the 17th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '90, pages 177–185, New York, NY, USA, 1990. ACM.

[Bry92] Steve Bryson. Virtual spacetime: An environment for the visualization of curved spacetimes via geodesic flows. In *Proceedings of the 3rd Conference on Visualization '92*, VIS '92, pages 291–298, Los Alamitos, CA, USA, 1992. IEEE Computer Society Press.

[Bry96] Steve Bryson. Virtual reality in scientific visualization. *Commun. ACM*, 39(5):62–71, May 1996.

[BSH+05] Jeremy N. Bailenson, Kim Swinth, Crystal Hoyt, Susan Persky, Alex Dimov, and Jim Blascovich. The independent and interactive effects of embodied-agent appearance and behavior on self-report, cognitive, and behavioral markers of copresence in immersive virtual environments. *Presence: Teleoperators and Virtual Environments*, 14(4):379–393, aug 2005.

[But16] J.C. Butcher. *Numerical Methods for Ordinary Differential Equations*. Wiley, 2016.

[BY05] Jeremy N Bailenson and Nick Yee. Digital chameleons: automatic assimilation of nonverbal gestures in immersive virtual environments. *Psychol Sci*, 16(10):814–819, oct 2005.

[Car82] Carroll. The adventure of getting to know a computer. *Computer*, 15(11):49–58, nov 1982.

[CB16] James J. Cummings and Jeremy N. Bailenson. How immersive is enough? a meta-analysis of the effect of immersive technology on user presence. *Media Psychol*, 19(2):272–309, apr 2016.

[CKR01] Vanessa Stevens Colella, Eric Klopfer, and Mitchel Resnick. *Adventures in Modeling: Exploring Complex, Dynamic Systems with Starlogo*. Teachers College Press, New York, NY, USA, 1st edition, 2001.

[DBB+03] Yehudit Judy Dori, John Belcher, Mark Bessette, Michael Danziger, Andrew McKinney, and Erin Hult. Technology for active learning. *Materials Today*, 6(12):44 – 49, 2003.

[DC96] James N Davidson and Dace A Campbell. Collaborative design in virtual space - greenspace ii: A shared environment for architectural design review. In Patricia McIntosh and Filiz Ozel, editors, *Design Computation: Collaboration, Reasoning, Pedagogy: Acadia Conference Proceedings*, pages 165–179, Tucson, USA, oct 1996. ACADIA.

[Ded95] Chris Dede. The evolution of constructivist learning environments: Immersion in distributed, virtual worlds. *Educational technology*, 35(5):46–52, 1995.

[Ded09] Chris Dede. Immersive interfaces for engagement and learning. *Science*, 323(5910):66–69, 2009.

[Dew33]   John Dewey. *How we think: A restatement of the reflective thinking to the educative process.* Heath, 1933.

[dJLZ13]   Ton de Jong, Marcia C. Linn, and Zacharias C. Zacharia. Physical and virtual laboratories in science and engineering education. *Science*, 340(6130):305–308, 2013.

[Dou04]   P. Dourish. *Where the Action is: The Foundations of Embodied Interaction.* Bradford books. Bradford Books, 2004.

[DRW$^+$11] Sigrun Doberenz, Walton T Roth, Eileen Wollburg, Nina I Maslowski, and Sunyoung Kim. Methodological considerations in ambulatory skin conductance monitoring. *International Journal of Psychophysiology*, 80(2):87–95, 2011.

[DSL96]   C. Dede, M. C. Salzman, and R. Bowen Loftin. Sciencespace: virtual realities for learning complex and abstract scientific concepts. In *Proceedings of the IEEE 1996 Virtual Reality Annual International Symposium*, pages 246–252, 271, Mar 1996.

[FAJ$^+$15] Jesse Fox, Sun Joo (Grace) Ahn, Joris H. Janssen, Leo Yeykelis, Kathryn Y. Segovia, and Jeremy N. Bailenson. Avatars versus agents: A meta-analysis quantifying the effect of agency on social influence. *Human–Computer Interaction*, 30(5):401–432, sep 2015.

[FI87]   Thomas A Furness III. Designing in virtual space. 1987.

[GCM17]   Scott W Greenwald, Wiley Corning, and Pattie Maes. Multi-user framework for collaboration and co-creation in virtual reality. In *12th International Conference on Computer Supported Collaborative Learning*, 2017.

[GP92]   Mary L Gick and Katherine Paterson. Do contrasting examples facilitate schema acquisition and analogical transfer? *Canadian Journal of Psychology/Revue canadienne de psychologie*, 46(4):539, 1992.

[GSV$^+$03] Maia Garau, Mel Slater, Vinoba Vinayagamoorthy, Andrea Brogni, Anthony Steed, and M. Angela Sasse. The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment. In *Proceedings of the conference on Human factors in computing systems - CHI '03*, page 529, New York, New York, USA, apr 2003. ACM Press.

[GWFM17]   Scott W. Greenwald, Zhangyuan Wang, Markus Funk, and Pattie Maes. *Investigating Social Presence and Communication with Embodied Avatars in Room-Scale Virtual Reality*, pages 75–90. Springer International Publishing, Cham, 2017.

[Haa84]   M. W. Haas. Visually coupled systems as simulation devices. *J Aircr*, 21(8):639–640, aug 1984.

[HPGK94]   Ken Hinckley, Randy Pausch, John C. Goble, and Neal F. Kassell. A survey of design issues in spatial input. In *Proceedings of the 7th Annual ACM Symposium on User Interface Software and Technology*, UIST '94, pages 213–222, New York, NY, USA, 1994. ACM.

[HS92]   Jim Hollan and Scott Stornetta. Beyond being there. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '92, pages 119–125,

New York, NY, USA, 1992. ACM.

[HSZ17]   Ryan Hubbard, Aldis Sipolins, and Lin Zhou. Enhancing learning through virtual reality and neurofeedback: A first step. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, LAK '17, pages 398–403, New York, NY, USA, 2017. ACM.

[IT94]   Ellen A. Isaacs and John C. Tang. What video can and cannot do for collaboration: A case study. *Multimedia Systems*, 2(2):63–73, aug 1994.

[Kay72]   Alan C. Kay. A personal computer for children of all ages. In *Proceedings of the ACM Annual Conference - Volume 1*, ACM '72, New York, NY, USA, 1972. ACM.

[KB15]   Cody O. Karutz and Jeremy N. Bailenson. *Immersive Virtual Environments and the Classrooms of Tomorrow*, pages 290–310. John Wiley & Sons, Ltd, 2015.

[KGS12]   Konstantina Kilteni, Raphaela Groten, and Mel Slater. The sense of embodiment in virtual reality. *Presence: Teleoperators and Virtual Environments*, 21(4):373–387, nov 2012.

[Kla02]   Frank Klassner. A case study of lego mindstorms'™ suitability for artificial intelligence and robotics courses at the college level. *SIGCSE Bull.*, 34(1):8–12, February 2002.

[KLSB14]   Seungwon Kim, Gun Lee, Nobuchika Sakata, and Mark Billinghurst. Improving co-presence with augmented visual communication cues for sharing experience through video conference. In *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 83–92. IEEE, sep 2014.

[Koc77]   Dean F Kocian. A visually-coupled airborne systems simulator (vcass)-an approach to visual simulation. Technical report, AIR FORCE AEROSPACE MEDICAL RESEARCH LAB WRIGHT-PATTERSON AFB OH, 1977.

[Kol84]   D.A. Kolb. *Experiential Learning: Experience as the Source of Learning and Development*. Prentice-Hall, 1984.

[LB11]   L. Lin and C. Bigenho. Note-taking and memory in different media environments. *Computers in the Schools*, 28(3):200–216, 2011.

[LB15]   L. Lin and C. Bigenho. Learning in technology-immersive learning environments. In *The Wiley Blackwell Handbook of Psychology, Technology and Society*, pages 420–425. 2015.

[LBB99]   JM Loomis, JJ Blascovich, and AC Beall. Immersive virtual environment technology as a basic research tool in psychology. *&*, 1999.

[Lom14]   E. Lombardo. Study of an interactive and total immersive device with a personal 3d viewer and its effects on the explicit long-term memories of the subjects. In *Designing and Developing Virtual and Augmented Environments*, volume 8525 of *LNCS*, pages 75–84. 2014.

[LW91]   Jean Lave and Etienne Wenger. *Situated Learning: Legitimate Peripheral Participation*. Cambridge University Press, reprint edition, 1991.

[Mal82]   Thomas W. Malone. Heuristics for designing enjoyable user interfaces: Lessons from computer games. In *Proceedings of the 1982 conference on Human factors in computing systems - CHI '82*, pages 63–68, New York, New York, USA, mar 1982. ACM Press.

[MBK+04]   J. Maloney, L. Burd, Y. Kafai, N. Rusk, B. Silverman, and M. Resnick. Scratch: a sneak preview [education]. In *Proceedings. Second International Conference on Creating, Connecting and Collaborating through Computing, 2004.*, pages 104–109, Jan 2004.

[MC08]   Bradford Z Mahon and Alfonso Caramazza. A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *J Physiol Paris*, 102(1-3):59–70, may 2008.

[MF67]   Marshall McLuhan and Quentin Fiore. The medium is the message. *New York*, 123:126–128, 1967.

[MGSGN13] Manolis Mavrikis, Sergio Gutierrez-Santos, Eirini Geraniou, and Richard Noss. Design requirements, student perception indicators and validation metrics for intelligent exploratory learning environments. *Pers Ubiquitous Comput*, 17(8):1605–1620, dec 2013.

[MO14]   P.A. Mueller and D.M. Oppenheimer. The pen is mightier than the keyboard: advantages of longhand over laptop note taking. *Psychol Sci*, 25(6):1159–1168, 2014.

[NDJ93]   Melanie Njoo and Ton De Jong. Exploratory learning with a computer simulation for control theory: Learning processes and instructional support. *J Res Sci Teach*, 30(8):821–844, oct 1993.

[New13]   Nora S Newcombe. Seeing relationships: Using spatial thinking to teach science, mathematics, and social studies. *American Educator*, 2013.

[OMIM94]   Ken-Ichi Okada, Fumihiko Maeda, Yusuke Ichikawaa, and Yutaka Matsushita. Multiparty videoconferencing at virtual social distance. In *Proceedings of the 1994 ACM conference on Computer supported cooperative work - CSCW '94*, pages 385–393, New York, New York, USA, oct 1994. ACM Press.

[Pan10]   Veronica S Pantelidis. Reasons to use virtual reality in education and training courses and a model to determine when to use virtual reality. *Themes in Science and Technology Education*, oct 2010.

[Pap72]   Seymour Papert. Teaching children to be mathematicians versus teaching about mathematics. *International Journal of Mathematical Education in Science and Technology*, 3(3):249–262, 1972.

[PAT99]   Gorman PJ, Meier AH, and Krummel TM. Simulation and virtual reality in surgical education: Real or unreal? *Archives of Surgery*, 134(11):1203–1208, 1999.

[PH91]   S Papert and I Harel. Situating constructionism. *Constructionism*, 36(2):1–11, 1991.

[Pia52]   J. Piaget. *La naissance de l'intelligence chez l'enfant*. Norton library. International Universities Press, 1952.

[POK05]   A. Piolat, T. Olive, and R.T. Kellogg. Cognitive effort during note taking. *Appl Cogn Psychol*, 19(3):291–312, 2005.

[PR09]   Ruth Paradise and Barbara Rogoff. Side by side: learning by observing and pitching in. *Ethos*, 37(1):102–138, mar 2009.

[PTW98]   I. Poupyrev, N. Tomokazu, and S. Weghorst. Virtual notepad: handwriting in immersive vr. In *Proceedings. IEEE 1998 Virtual Reality Annual International Symposium (Cat. No.98CB36180)*, pages 126–132. IEEE Comput. Soc, 1998.

[Pur13]   E.M. Purcell. *Electricity and Magnetism*. Electricity and Magnetism. Cambridge University Press, 2013.

[RAR15]   Martina A. Rau, Vincent Aleven, and Nikol Rummel. Successful learning with multiple graphical representations and self-explanation prompts. *J Educ Psychol*, 107(1):30–46, 2015.

[Reg97]   J. Wesley Regian. *Virtual Reality for Training: Evaluating Transfer*, pages 31–40. Springer US, Boston, MA, 1997.

[Rie96]   John Rieman. A field study of exploratory learning strategies. *ACM Trans. Comput.-Hum. Interact.*, 3(3):189–218, September 1996.

[RMF15a]   Martina A. Rau, Joseph E. Michaelis, and Natalie Fay. Connection making between multiple graphical representations: A multi-methods approach for domain-specific grounding of an intelligent tutoring system for chemistry. *Comput Educ*, 82:460–485, mar 2015.

[RMF15b]   Martina A. Rau, Joseph E. Michaelis, and Natalie Fay. Connection making between multiple graphical representations: A multi-methods approach for domain-specific grounding of an intelligent tutoring system for chemistry. *Computers & Education*, 82:460 – 485, 2015.

[RMG+93]   Barbara Rogoff, Jayanthi Mistry, Artin Göncü, Christine Mosier, Pablo Chavajay, Shirley Brice Heath, and Artin Goncu. Guided participation in cultural activity by toddlers and caregivers. *Monogr Soc Res Child Dev*, 58(8):i, 1993.

[RMN+05]   Mitchel Resnick, Brad Myers, Kumiyo Nakakoji, Ben Shneiderman, Randy Pausch, and Mike Eisenberg. Design principles for tools to support creative thinking. 20, 01 2005.

[ROS06]   Maria Roussou, Martin Oliver, and Mel Slater. The virtual playground: an educational virtual reality environment for evaluating interactivity and conceptual learning. *Virtual Reality*, 10(3):227–240, Dec 2006.

[RPA+03]   Barbara Rogoff, Ruth Paradise, Rebeca Mejía Arauz, Maricela Correa-Chavez, and Cathy Angelillo. Firsthand learning through intent participation. *Annu Rev Psychol*, 54:175–203, 2003.

[RSM92]   J. Wesley Regian, Wayne L. Shebilske, and John M. Monk. Virtual reality: An instructional medium for visual-spatial tasks. *Journal of Communication*, 42(4):136–149, 1992.

[RY94]   J. Wesley Regian and Robert M. Yadrick. Assessment of configurational knowledge of naturally- and artificially-acquired large-scale space. *Journal of Environmental Psychology*, 14(3):211 – 223, 1994.

[SAK11]   J Sweller, P Ayres, and S Kalyuga. *Cognitive load theory.* 2011.

[SB07]   Philip Schuchardt and Doug A. Bowman. The benefits of immersion for spatial understanding of complex underground cave systems. In *Proceedings of the 2007 ACM Symposium on Virtual Reality Software and Technology*, VRST '07, pages 121–124, New York, NY, USA, 2007. ACM.

[SC13]   R.J. Seidel and P.R. Chatelier. *Virtual Reality, Training's Future?: Perspectives on Virtual Reality and Related Emerging Technologies.* Defense Research Series. Springer US, 2013.

[SCH12]   Christopher Stapleton, Creative Venture Catalyst, and Atsusi Hirumi. Designing interplay learning landscapes to evoke emotions, spark the imagination. 2012.

[SHA15]   Pooja G. Sidney, Shanta Hattikudur, and Martha W. Alibali. How do contrasting cases and self-explanation promote learning? evidence from fraction division. *Learning and Instruction*, 40:29 – 38, 2015.

[Shn83]   B. Shneiderman. Direct manipulation: A step beyond programming languages. *Computer*, 16(8):57–69, Aug 1983.

[SKJGZ$^+$14] Jinseop S Kim, Matthew J Greene, Aleksandar Zlateski, Kisuk Lee, Mark Richardson, Srinivas Turaga, Michael Purcaro, Matthew Balkam, Amy Robinson, Bardia Behabadi, Michael Campos, Winfried Denk, Hyunjune Seung, and Rafikh Shaikh. Space-time wiring specificity supports direction selectivity in the retina. *Nature*, 509, May 2014.

[Sla99]   Mel Slater. Measuring presence: A response to the witmer and singer presence questionnaire. *Presence: Teleoperators and Virtual Environments*, 8(5):560–565, oct 1999.

[SSV14]   M. Slater and M. V. Sanchez-Vives. Transcending the self in immersive virtual reality. *Computer*, 47(7):24–30, July 2014.

[SUS94]   M Slater, M Usoh, and A Steed. Depth of presence in virtual environments. *: Teleoperators & Virtual Environments*, 1994.

[TAF86]   III Thomas A. Furness. The super cockpit and its human factors challenges. *Proceedings of the Human Factors Society Annual Meeting*, 30(1):48–52, 1986.

[VK78]   Steven G. Vandenberg and Allan R. Kuse. Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills*, 47(2):599–604, 1978.

[Wil02a]   Margaret Wilson. Six views of embodied cognition. *Psychon Bull Rev*, 9(4):625–636, dec 2002.

[Wil02b]   Margaret Wilson. Six views of embodied cognition. *Psychonomic Bulletin & Review*, 9(4):625–636, 2002.