

Making the Abstract Physical: Toward Improving Binary Reverse Engineering via Embodied Immersion

by

Dennis G. Brown, Jr.

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
August 9, 2025

Keywords: virtual reality, cognition, binary reverse engineering, sensemaking, program comprehension

Copyright 2025 by Dennis G. Brown, Jr.

Approved by

Samuel Mulder, Chair, Associate Research Professor of Computer Science and Software Engineering
N. Hari Narayanan, Department Chair and John H. and Gail Watson Professor of Computer Science and Software Engineering
Cheryl D. Seals, Charles W. Barkley Professor of Computer Science and Software Engineering
Mark Yampolskiy, Associate Professor of Computer Science and Software Engineering

Abstract

Performing reverse engineering (RE) to determine precisely what a piece of software does is vital to tasks such as securing networks, mitigating malware, and maintaining legacy software. Much software is distributed as binary executable programs, which are particularly difficult for humans to comprehend because the compilation process is a one-way transformation from context-rich source code to a highly-optimized binary program. Our central problem is that binary RE is a highly-specialized skill that requires extensive training and experience. Additionally, the RE process requires a human-in-the-loop because the compound uncertainties introduced in disassembling and decompiling a binary program prevent a fully-automated solution. This work postulates and tests methods to improve the effectiveness of the human-machine joint cognitive system performing sensemaking in the context of binary RE, particularly, leveraging affordances of immersive virtual reality (VR) that exploit facets of cognition typically underutilized in the task of binary RE.

In tackling this problem, we followed a hybrid human-centered interaction design process combining Design Thinking (DT) with Cognitive Systems Engineering (CSE). Within our discovery phase, we performed a thorough interdisciplinary survey providing the theoretical basis for augmented the RE process with immersive VR. In our definition phase, we prioritized the identified affordances in VR into an initial set for the development phase, launching into multiple iterations of build and test of our VR system, Cognitive Binary Reverse Engineering (CogBRE), leveraging feedback from RE practitioners in each iteration.

Based on informal feedback from these iterations indicating that practitioners value the ability to place and organize code fragments and flow graphs in the expanse of VR as they form an understanding of a binary program, we designed and executed a formal user study. Using a between-subjects design, we compared CogBRE (in two VR configurations) to a traditional desktop interface across several metrics, including performance, cognitive load, usability, and

user experience. While task accuracy was statistically equivalent across conditions, participants in VR reported significantly lower cognitive load and described meaningful advantages in spatial organization and contextual reasoning.

Recognizing the rise of large language models (LLMs) and spurred by early feedback that users sought novel graph visualizations, we conducted a pilot study using an LLM to generate 3D function call graphs based on self-directed interrogation of disassembled binaries. Evaluators found that the LLM-generated visualizations were often correct, interpretable, and helpful, suggesting future potential for LLMs as collaborative agents in immersive environments.

This work contributes a novel VR system to augment the binary RE process, empirical evidence that embodied spatial interaction can reduce subjective cognitive load, early evidence that LLMs can serve as visualization and reasoning partners, and a generalizable research framework for applying embodied interaction to other cognitively demanding technical domains.

Acknowledgments

The journey through a doctoral program is never a solo effort and I need to thank those who supported me in various ways:

My advisor, Dr. Samuel Mulder, for taking in a wayward graduate student with a mutual interest in Virtual Reality, and Dr. Daniel Tauritz for helping make the connection.

My committee members, Dr. Hari Narayanan, Dr. Cheryl Seals, Dr. Mark Yampolskiy, and University Reader Dr. Davide Guzzetti, for your service and feedback in this process.

Dr. Xiao Qin and Clint Lovelace for their kind and competent guidance throughout my experience.

The Auburn Program Understanding Lab members, for entertaining my exploits, and especially Kevan Baker, Julian Bauer, and Luke Wittbrodt for providing hands-on support in building and evaluating the software system providing the foundation of this work.

The United States Air Force Science & Engineering Career Field Team's Science, Technology, Engineering, Mathematics + Management (STEM+M) program, for financially supporting my education.

My parents, Dennis and Sandra Brown, for ingraining in me that education is worth sacrifice, and my sister, Dr. Serena Brown, for raising the bar.

My children, Jessica (Wrenley), Alison, Rachel, and Zachary, for dealing with my limited free time over the last five years, and my fuzzy children for keeping me company during long coding and writing sessions.

My wife, best friend, and soulmate Lisa Brown, for her unwavering and invaluable support on this journey.

Table of Contents

Abstract	ii
Acknowledgments	iv
List of Tables	ix
List of Figures	x
1 Introduction	1
1.1 Research Problem	1
1.2 Research Questions	2
1.3 Research Approach	3
2 Literature Survey	5
2.1 Why is Binary RE Difficult?	6
2.2 Methodology	7
2.2.1 Methodology for Identifying Characteristics of Mental/Cognitive Models of Sensemaking in Binary RE (Section 2.3)	7
2.2.2 Methodology for Reviewing Underlying Cognitive Theory (Section 2.4)	8
2.2.3 Methodology for Reviewing Cognitive Augmentation Using Visualization and Immersion (Section 2.5)	9
2.3 Characteristics of Mental/Cognitive Models of Sensemaking in Binary RE	9
2.3.1 Discovering Mental/Cognitive Models of binary RE	10
2.3.2 Differences between Novices and Experts in Performing binary RE	15
2.3.3 Section Summary	16

2.4	Underlying Cognitive Theory and Applications to Binary RE	17
2.4.1	External and Embodied Cognition	17
2.4.2	Cognitive Load Theory	19
2.4.3	Applications to Program Comprehension	20
2.4.4	Section Summary	21
2.5	Cognitive Augmentation of Sensemaking Using Visualization and Immersion .	22
2.5.1	Leveraging Affordances of Immersive Technologies for Analytic Tasks	22
2.5.2	Visualization Applied to Software Reverse Engineering	27
2.5.3	Section Summary	31
2.6	Synthesis	32
2.6.1	Overview of Elements and Connecting Threads	32
2.6.2	Primary Themes	35
2.6.3	Section Summary	39
2.7	Conclusion	40
2.7.1	Threats to Validity	41
3	Design Process and System Architecture	42
3.1	Design Thinking and Cognitive Systems Engineering	42
3.1.1	Discovery Phase	43
3.1.2	Definition Phase	45
3.1.3	Development → Delivery Phase Iterations	46
3.2	System Architecture	46
3.2.1	Oxide	48
3.2.2	Nexus	48
3.2.3	VR Client	49
3.3	Iteration 1: Minimally Viable Product	51

3.3.1	Feature Implementation	51
3.3.2	Feedback	52
3.4	Iteration 2	56
3.4.1	Feature Implementation	56
3.4.2	Feedback	57
3.5	Iteration 3	64
3.6	Conclusion	66
4	User Study	67
4.1	Study Design	67
4.1.1	Hypotheses	68
4.1.2	Conditions	69
4.1.3	Participants	73
4.1.4	RE Challenge	75
4.1.5	Data Collection	79
4.1.6	Procedure	81
4.1.7	Approval	83
4.1.8	Threats to Validity	83
4.2	Data and Analysis	85
4.2.1	Data Overview	85
4.2.2	Analysis	101
4.3	Conclusion and Future Work	112
5	An Experiment with AI	114
5.1	Related Work and Motivation	116
5.2	System Overview: LLM-Augmented CogBRE	117
5.3	Evaluation Design	118

5.4	Prompting Methodology	120
5.5	Evaluation Results and Analysis	122
5.6	Conclusion	127
6	Conclusion	128
6.1	Summary of Contributions	129
6.2	Interpretation of Findings	130
6.3	Limitations	131
6.4	Future Work	131
6.5	Closing Remarks	132
	References	133
	Appendices	147
A	Reverse Engineering Challenge source code listing	148
B	User Study Playbook	152
C	Institutional Review Board Package	167

List of Tables

2.1	Elements of (A) cognitive models of binary RE, (B) concepts of cognitive theory, and (C) VR affordances/tools/techniques	34
4.1	Summary of the data we collected.	81
4.2	Statistical overview of self-reported skill levels in relevant areas (on scale 1-5).	87
4.3	Statistical overview of correctness of RE Challenge responses.	89
4.4	Statistical overview of duration of RE Challenge responses (seconds).	90
4.5	Statistical overview of TLX responses (on scale 0-100).	92
4.6	Statistical overview of SUS responses to positive-sentiment items (on scale 1-5).	94
4.7	Statistical overview of SUS responses to negative-sentiment items (on scale 1-5).	95
4.8	Statistical overview of responses to first set of Presence items (on scale 1-7).	96
4.9	Statistical overview of responses to second set of Presence items (on scale 1-7).	97
4.10	Statistical overview of responses to third set of Presence items (on scale 1-7).	97
4.11	Text of Presence questionnaire items for Tables 4.8 to 4.10 and Figures 4.16 to 4.17.	98
4.12	Statistical overview of ratings of affordances in VR (on scale 1-5).	100
4.13	Statistical overview of volumes (VR Enhanced) and areas (VR Basic) occupied by graphs and slates.	102
4.14	Kruskal-Wallis H-stat and p-value for the three conditions for each pre-survey experience rating.	105
5.1	Objective measures across 40 graphs.	123
5.2	Mean and Coefficient of Variation for Subjective and Objective Ratings.	124

List of Figures

1.1	Compiling source code to a assembly code is a one-way process discarding contextual information that helps humans understand the code.	2
2.1	RE process per Shneiderman and Mayer [1], adapted and reframed as the central process in program comprehension / reverse engineering	11
2.2	RE process based on sensemaking; adapted from Bryant et al. [2] and reframed as the central process in program comprehension / reverse engineering	12
2.3	RE process based on abductive iteration; adapted from Dudenhofer [3] and reframed as the central process in program comprehension / reverse engineering	14
2.4	RE process in three phases, adapted from Votipka et al. [4] and reframed as the central process in program comprehension / reverse engineering	15
2.5	Threads connecting elements of (A) cognitive models of binary RE, (B) concepts of cognitive theory, and (C) VR affordances/tools/techniques	33
2.6	Primary themes for analysis with most closely-related elements; highlighted area indicates highest-priority VR affordances	36
3.1	Double diamond.	43
3.2	Hybrid process detail.	44
3.3	Empathy map.	45
3.4	CogBRE Platform Architecture	47
3.5	The VR Client, as a Unity application, consists of modular “managers” extending the MonoBehavior base class that run in the main thread; some of them spawn threads to accomplish long-running tasks with minimal impact to the main thread, improving interactivity and usability.	50
3.6	Iteration 1 (MVP) Features	52
3.7	Iteration 1 screenshot: the main menu and a strings slate; text scrolling in slates was not yet implemented.	53
3.8	Iteration 1 screenshot: a call graph and informational slates.	53

3.9	Iteration 1 screenshot: function disassemblies and decompilations.	54
3.10	Iteration 1 screenshot: decompilation slates and control flow graph.	54
3.11	Iteration 2 Features: Updates are in bold (note taking and signalling).	57
3.12	Iteration 2 screenshot: the main menu in the new color theme and “office” surroundings.	58
3.13	Iteration 2 screenshot: a slate, in the new color theme and with the “copy” button added, showing strings extracted from a binary.	58
3.14	Iteration 2 screenshot: the radial menu allows users to enable and disable the main menu, notepad, and magnifying glass from the touchpad of the VR controller.	59
3.15	Iteration 2 screenshot: as an alternative to panning and zooming a large graph, the user can use the magnifying glass.	59
3.16	Iteration 2 screenshot: the new notepad and on-screen keyboard; in this iteration, the keyboard was severable from the notepad.	60
3.17	Iteration 2 screenshot: call graph functions with Capa findings are shown with visual distinction.	60
3.18	Iteration 2 screenshot: large graphs exceed space available in the new environment.	63
3.19	Iteration 3 screenshot: a large call graph (for Windows XP REGEDIT.EXE) fits easily in this environment, though the user still needs to pan/zoom/magnify as needed.	64
3.20	Iteration 3 screenshot: main menu and a text slate in the new open environment and text improvements.	65
3.21	Iteration 3 screenshot: a call graph for a simple binary with a control flow graph for one of its functions; darkened area indicates the grabbable area of the graph.	65
3.22	Iteration 3 screenshot: updated notepad with attached keyboard that provides direct control of the notepad text area as most users expect.	66
4.1	Web Client added to the CogBRE architecture.	70
4.2	Web Client running in a browser supporting a user session.	71
4.3	VR Basic representative screenshot showing, left-to-right, the notepad, question control, main menu, and four text slates in the semi-circular layout.	72

4.4	VR Enhanced representative screenshot showing the default components (notepad, question control, main menu) and additional items spawned and arranged by the user: four text slates, a function call graph, and a control flow graph.	73
4.5	Call graph for the “v8” binary used for the reverse engineering challenge	76
4.6	Pre-survey ratings box plot.	87
4.7	Pre-survey ratings radar plot.	88
4.8	RE Challenge Question Correctness box plot.	89
4.9	RE Challenge Question Correctness radar plot.	90
4.10	RE Challenge Question Duration box plot.	91
4.11	RE Challenge Question Duration radar plot.	91
4.12	TLX box plot.	92
4.13	TLX radar plot.	93
4.14	SUS positive-sentiment radar plot.	94
4.15	SUS negative-sentiment radar plot.	95
4.16	Presence box plot.	97
4.17	Presence radar plot.	99
4.18	Affordance ratings box plot.	100
4.19	Volume used per question for each participant in the VR Enhanced condition; two participants were excluded due to being outliers.	102
4.20	Area used per question for each participant in the VR Basic condition.	103
4.21	Average number of items open per question.	103
5.1	Call graph designed by an LLM after performing self-directed discovery of the v8 binary program employed in the user study discussed in Chapter 4. Note how it arranged the chains of functions (part1a..d, etc.) into rows and gave them a specific shape and color.	115
5.2	Call graph designed by an LLM after performing self-directed discovery of a binary program simulating a remote-controlled bot. Note overall layout in hemispherical shape with higher-level functions in the center radiating out to groups of related functions and system calls on the periphery.	115
5.3	CogBRE System Architecture with LLM and MCP.	117

5.4	Basic 2D call graph for v11 (without external calls).	125
5.5	Highest-rated call graph for v11 (subjective): High guidance, GPT-4.1.	125
5.6	Lowest-rated call graph for v11 (subjective): High guidance, GPT-4.1.	125
5.7	Basic 2D call graph for hexdump (without external calls).	126
5.8	Highest-rated call graph for hexdump (subjective): High guidance, o4-mini. . .	126
5.9	Lowest-rated call graph for hexdump (subjective): High guidance, o4-mini. . .	126

Chapter 1

Introduction

1.1 Research Problem

Understanding the purpose and effects of an existing software program, in absence of significant documentation about that software, is critical to many tasks such as maintaining and improving that software and mitigating potential malicious impacts of that software. This process is commonly called *reverse engineering* (RE), which is the term we will use in this paper, but also called *program comprehension* or *program understanding*. Comprehending non-trivial *binary* programs is quite difficult [5]. Most binary programs are a result of compiling source code, optimization, and linking where important context about the program is lost in every step of a multi-step process, as illustrated by Figure 1.1. Many tools and approaches have been developed over time to analyze, disassemble, and decompile binary programs, including adopting artificial intelligence [6] [7] [8] [9] [10]. Common RE approaches somewhat seek to reverse the compilation process, but the loss at each step introduces significant uncertainty. This uncertainty thwarts fully automating binary RE—the process requires a human in the loop, one with rare knowledge and experience.

One key to improving this human-centric process is understanding how a human operates within the process. Cognitive Systems Engineering (CSE) provides a framework for building that understanding, in which one considers holistically the combination of the operator and the

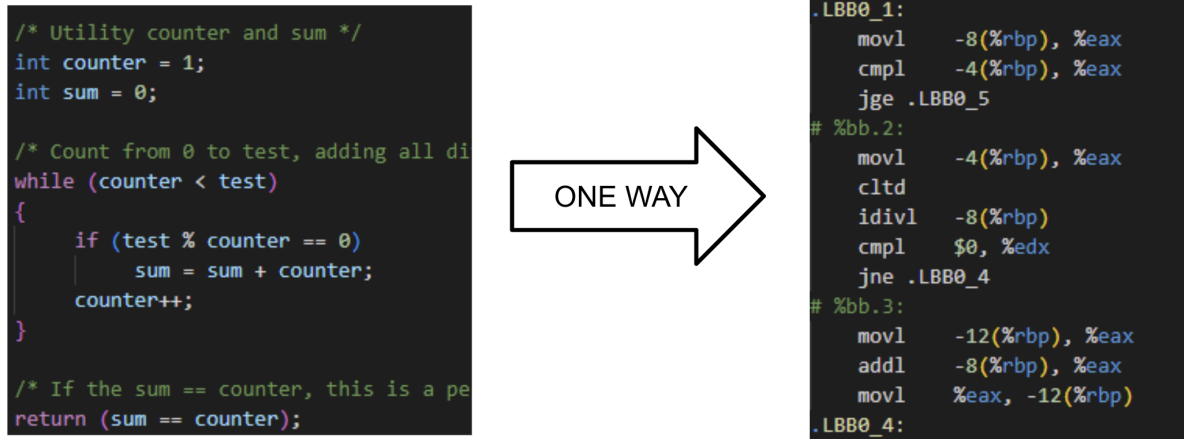


Figure 1.1: Compiling source code to a assembly code is a one-way process discarding contextual information that helps humans understand the code.

machine—the joint cognitive system—and how to make that joint system more effective by reducing cognitive complexity for the user [11]. In this approach, all work performed by humans is a form of cognitive work. Much prior research into the cognitive processes involved in performing binary RE casts the problem as a sensemaking task [2], and that work has developed cognitive models of the process [4] [12] [13] and applied Cognitive Load Theory (CLT) [14].

Employing embodied cognition is posed as one method of improving cognitive processes [15], and immersive Virtual Reality (VR) is an accessible means of providing embodied user experiences, especially for inherently abstract tasks like binary RE. In particular, Immersive Analytics (IA) [16] [17] [18] is an emergent and promising area of research that uses innovative display and interaction techniques to aid humans in the understanding of complex analytic problems by providing an egocentric experience. IA may provide the conceptual building blocks for significantly improving the task of binary RE.

1.2 Research Questions

This dissertation investigates how we might reduce the cognitive complexity of binary RE tasks via the employment of immersive interaction in order to improve the overall outcomes of the joint cognitive system. The specific research questions are as follows:

- RQ1. What are common and significant characteristics of mental/cognitive models of sense-making employed by binary RE practitioners?
- RQ2. What are underlying elements of cognitive theory bridging sensemaking in binary RE and immersive sensemaking?
- RQ3. How have the affordances of immersive virtual reality been employed to improve cognition and sensemaking in analytic tasks?
- RQ4. How can we use these findings to effect improvements in the practice of binary RE?
- RQ5. What design and architectural features are required in an immersive VR system to support static binary RE?
- RQ6. How do binary RE practitioners employ the affordances of embodied immersion in this system?
- RQ7. Does this system improve outcomes in static binary RE tasks?
- RQ8. What are ways an AI-based agent could improve the user experience?

Our general hypothesis is that applying techniques of immersive embodied interaction that are traceable to cognitive elements of the binary RE process can reduce the cognitive complexity of binary RE, improving outcomes in specific situations and informing broader applications of this approach.

1.3 Research Approach

To investigate the potential of immersive virtual reality to improve the practice of binary RE, we followed a rigorous, interdisciplinary research approach grounded in Cognitive Systems Engineering (CSE) and Design Thinking (DT). These frameworks support both the theoretical and practical exploration of our central hypothesis: that embodied immersion can reduce the cognitive complexity of static binary RE tasks, thereby improving user experience and task performance.

In Chapter 2, we synthesize findings from diverse fields (Reverse Engineering and Program Comprehension; Cognition and Cognitive Systems; Visualization; Virtual Reality; and Immersive Analytics) to address RQs 1 through 4. This literature survey establishes a foundation for understanding the cognitive demands of binary RE and the potential of immersive affordances to augment human sensemaking.

Drawing from this synthesis, we designed and implemented a novel immersive RE platform. Chapter 3 details our process developed as a hybrid of CSE and DT. This process starts with identifying key affordances and user needs, moves to iterative prototyping and refinement, and results in a working system that integrates standard RE tool outputs into an embodied, spatial interface. This work addresses RQ5.

Chapter 4 describes a between-subjects user study comparing desktop and VR-based RE experiences under controlled conditions. This study evaluates how participants engage with the system and whether certain affordances in VR lead to better subjective and objective outcomes. The results inform RQs 6 and 7, focusing on usability, cognitive load, and task performance.

Building on our core system, Chapter 5 explores a nascent form of affordance: the integration of a large language model (LLM) agent capable of synthesizing binary features and generating spatial visualizations. This preliminary study investigates the role of AI as a cognitive partner within immersive environments, addressing RQ8.

Finally, Chapter 6 synthesizes our findings and reflects on lessons learned, limitations, and pathways for future work. We propose next steps for advancing immersive systems for RE and consider broader applications of this research.

Throughout, our methods emphasize qualitative insight and quantitative rigor. Each phase builds logically on the prior, ensuring continuity between theory, design, implementation, evaluation, and exploration. In doing so, we aim not only to demonstrate proof-of-concept for immersive RE but also to contribute a replicable research framework for applying embodied interaction to other cognitively demanding technical domains.

Chapter 2

Literature Survey

This chapter forms part of our discovery process in the hybrid CSE-DT framework (as will be further described in Chapter 3). It reviews and ties together research contributions from multiple disciplines in order to answer RQs 1-4. These disciplines include Reverse Engineering and Program Comprehension; Cognition and Cognitive Systems; Visualization; Virtual Reality; and Immersive Analytics. The papers covered in this survey were identified through an iterative approach. We first performed exploratory ad-hoc searches of the literature using common scholarly search engines and indices looking for directly-related results in the primary topic area of using VR to improve binary RE. While we found very few close matches, we did find many adjacent works and through a clustering process, we identified the disciplines listed previously. We then iterated this search process until we filled out a set of recent and relevant research papers.

This chapter proceeds as follows. Section 2.1 dives deeper into what makes binary RE difficult. We describe our methodology in Section 2.2. In Section 2.3, we explore prior research in sensemaking and cognitive models of RE (primarily binary and adjacent forms of RE), and in Section 2.5, we explore how affordances in visualization and immersion have been applied toward sensemaking problems. We present Section 2.4, cognitive theory relevant to both of those sections, as a bridge between them. Finally, we synthesize the findings into high-level cross-discipline themes in Section 2.6.

2.1 Why is Binary RE Difficult?

First, as we mentioned in Chapter 1, this task is very unlikely to be fully automated largely due to the compound uncertainties in the reverse engineering process. More formally, Rice's Theorem [19] implies that deciding whether a given binary program contains any non-trivial property is formally undecidable. In practice, with intermingled code and data on modern architectures, there are many common code patterns that are unresolvable. Even with state-of-the-art tools (including Large Language Models), the process of disassembly introduces uncertainty and error. Each additional property built on top of disassembly is thus both undecidable and built on a representation including that uncertainty. Human input is required to overcome ambiguous situations and determine when a solution is good enough to satisfy the current objective, and even with human input our ability to answer questions about binary programs is severely limited.

Second, understanding what properties we care about in a program is hard; e.g., how do we know if there is a hidden trigger or a backdoor? Even given a source representation, these questions are challenging to answer, often reducing to a question of program correctness. Since programs rarely have a detailed formal specification, the *correct* behavior may be an open question. Reverse engineering a piece of software is an iterative, hypothesis-driven process, requiring the analyst to continually reformulate and ask questions relative to the current state of knowledge.

Finally, programs are arguably the most complex things ever engineered by humans [20]. In building complex software, the field of software engineering has developed many models for managing abstraction and interfaces between different components to allow many programmers to design a program as a team. Without that infrastructure in place, an entire non-trivial program is too complex for a human mind to fully understand. This is compounded by the fact that the compilation process strips helpful comments, structures, and metadata, leaving the analyst to try and recreate the layers of abstraction that the original authors used to manage the complexity.

2.2 Methodology

Given the cognitive nature of our inquiry, we chose to approach it as a Cognitive Systems Engineering (CSE) task. CSE broadly has three steps summarized as: (1) identify the problem areas of the task; (2) understand the circumstances of the identified problems, focusing on potential causes and influences of those problems; and (3) pursue practical solutions to the identified problem areas [11]. Based on our premise and use case, we formulated the following Research Questions (RQ) to answer in this paper:

- RQ1. What are common and significant characteristics of mental/cognitive models of sense-making employed by binary RE practitioners? (*corresponds to CSE step 1*)
- RQ2. What are underlying elements of cognitive theory bridging sensemaking in binary RE and immersive sensemaking? (*corresponds to CSE step 2*)
- RQ3. How have the affordances of immersive virtual reality been employed to improve cognition and sensemaking in analytic tasks? (*corresponds to CSE step 3*)
- RQ4. How can we use these findings to effect improvements in the practice of binary RE? (*corresponds to CSE step 3*)

Because our RQs cover multiple disciplines — Reverse Engineering and Program Comprehension; Cognition, Cognitive Systems, and Sensemaking; Visualization; VR; and IA — we divided our survey process into three parts to answer RQ1, RQ2, and RQ3 respectively in Sections 2.3, 2.4, and 2.5. We answer RQ4 with our synthesis in Section 2.6. In the remainder of this section, we explain the methodology we used to identify and select the papers we cover in each of survey sections of this paper. Note that this paper is not a systematic literature review.

2.2.1 Methodology for Identifying Characteristics of Mental/Cognitive Models of Sensemaking in Binary RE (Section 2.3)

The goal of the first part of our survey is to identify common elements of published cognitive models of sensemaking in binary RE (RQ1). We sought publications that present clear

cognitive or mental models of the sensemaking process for unfamiliar binary programs, and we adopted that criterion as our ultimate filter. In order to find appropriate publications for this niche topic, we employed a hybrid search strategy combining database searches and snowballing [21].

Our database searches employed common academic search engines (ACM Portal, IEEE Xplore, ScienceDirect, Scopus, and Web of Science) using the following search over titles, abstracts, and key words (modified as needed to match syntax requirements of each search engine): *((binary OR executable) AND ("reverse engineering" OR "program comprehension" OR "program understanding")) AND ("mental model*" OR "cognitive model*" OR sensemaking)*). These searches yielded a total 17 unique publications, and after review, we found that 6 met the filter criterion.

Next, we employed iterative backward (references) and forward (citations) snowballing starting with this core group of publications. This process revealed additional work in cognitive models for program understanding, not only for binary programs, but also *select* related work in the much larger and more mature body of knowledge covering source code understanding. Authors covering models of binary RE often referenced prior work in source-related program understanding to provide a foundation for their work, so we selectively included this foundational work to provide context.

2.2.2 Methodology for Reviewing Underlying Cognitive Theory (Section 2.4)

Our second goal is to extract and explore the fundamental cognitive phenomena related to the identified cognitive model elements and to immersive experiences. We took two approaches with this section of the survey. First, in order to provide context for fundamental cognitive concepts related to our scenario, we performed several targeted database searches of specific cognitive phenomena and summarized the results. In this way, Section 2.4 serves as a primer and as a bridge to the final survey section. Second, we performed additional targeted searches for publications discussing the identified phenomena in the specific context of software reverse engineering, which we summarize in the latter part of the section.

2.2.3 Methodology for Reviewing Cognitive Augmentation Using Visualization and Immersion (Section 2.5)

Our approach with this section parallels that of the section on cognitive model elements. We sought publications that specifically discuss visual and/or immersive affordances for sensemaking for abstract analytic problems, using that criterion as our filter. Again we combined database searches with snowballing. Using the search string (*immersi* OR "virtual reality"*) *AND sensemaking* we found 17 unique publications; 8 met our filter criterion. With that core group of papers, we reviewed additional publications identified through snowballing and included those that met the filter. In this process, we found several papers meeting our filter that specifically cover software engineering and RE, so we put them in their own subsection.

A note on Related Work: We did not find survey papers in our searches that cover our exact premise, thus we did not include a separate “related work” section in this paper. We did identify several survey papers following our methodology described above that overlap the constituent areas of our survey. We have included their contributions in the appropriate sections, clearly identifying them as survey papers.

Next, we will review mental/cognitive models used by analysts to overcome these difficulties when performing binary RE.

2.3 Characteristics of Mental/Cognitive Models of Sensemaking in Binary RE

In this section, we answer RQ1, “What are common and significant characteristics of mental/cognitive models of sensemaking employed by binary RE practitioners?” We examine publications investigating the cognitive process of binary RE. In the CSE framework referenced in Section 2.2, we identify the problem areas (step 1). Additionally, because binary RE shares many concepts and theory with source code RE, which has been more extensively covered in the literature, we will include some applicable work in source code RE in this section.

2.3.1 Discovering Mental/Cognitive Models of binary RE

The *mental model*, or cognitive model, concept originated with Craik [22], who asserted that humans build internal representations of the external world and use them to reason about why things happen and anticipate what might happen next. Johnson-Laird's [23] research in experimental psychology corroborated Craik's claims by finding that people employ mental models in their working memory to perform reasoning. [24] Cognitive models can be particularly powerful because they combine elements of both the problem domain or task area and the characteristics of human cognition. [25]

Shneiderman and Mayer [1] developed an early cognitive model of how programmers build up an internal semantic representation of a program based on evidence discovered in a series of experiments they performed. They asserted that the model must describe common programming tasks (composition, comprehension, debugging, modification, and learning) in terms of cognitive structures that programmers form in their memories and cognitive processes to use and build that knowledge. Knowledge is categorized as semantic (general concepts independent of language) or syntactic ("precise, detailed, and arbitrary" details, primarily language-specific). They cast the task of RE as subtasks of debugging, modification, and learning in which the programmer uses syntactic knowledge to form a multi-level internal semantic representation of the program. Lower levels (e.g., sequences of operations) and higher levels (what the program does) can be formed independently, and this encoding process is similar to Miller's chunking process [26], where smaller chunks of statements join to form larger chunks. The internal semantic representation of the program is strongly retained and widely accessible. Figure 2.1 illustrates their concept.

Introducing an iterative model, Brooks [27] envisioned the RE process as forming a mental tree that branches into increasingly complete hypotheses of how the code works and what it does. As a branch is found to be incorrect, the programmer cuts that branch and backtracks and may follow another higher-level branch or add a new one. This process is essentially abductive reasoning applied to code understanding, described by Weigand and Hartung [28] as a process

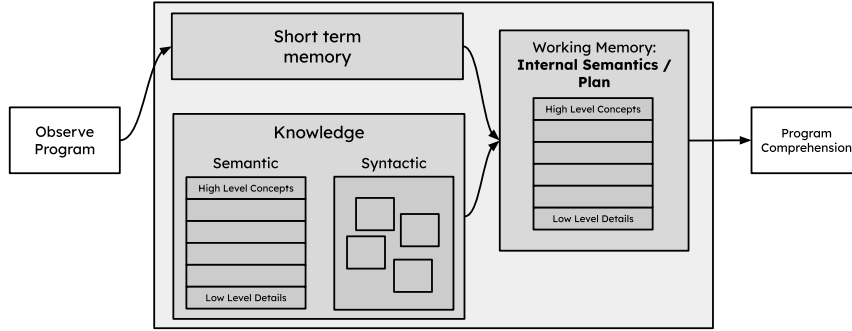


Figure 2.1: RE process per Shneiderman and Mayer [1], adapted and reframed as the central process in program comprehension / reverse engineering

of making observations, forming hypotheses, creating mental models of code that support the hypotheses, and searching for information to prove or disprove the hypotheses.

Zayour and Lethbridge [29] performed a cognitive analysis to identify the mental steps of source-code-based RE. In their analysis, the language was low-level proprietary code and not a typical high-level language on which RE is performed, increasing the applicability of their results to binary RE. They identified two primary cognitive difficulties of disorientation following recursions and disorientation in understanding the most relevant execution paths in the code. In response, they proposed two high level cognitive design requirements. The first is to minimize how many artifacts the engineer needs to keep in working memory by maintaining visual proximity between artifacts; linking new and existing information with meaningful encoding or chunking; and facilitating backtracking in execution paths. The second is to minimize fading of working memory by reducing the time artifacts need to be maintained in working memory and minimizing the number of steps between artifact acquisition. They used their findings to drive implementation of a RE assistive process in their DynaSee tool, which performs several filtering steps on program execution (remove redundancies, detect patterns, rank code routines) before visualizing trace patterns.

The Data-Frame Theory was established by Klein et al. [30] in which a frame—an explanatory structure such as a story, map, script, or plan that relates entities to other entities—is simultaneously fitted to discovered data and also drives the discovery of further data. The frame provides a foundation for understanding until flaws are detected. *Sensemaking* happens when

the frame is re-fitted to new evidence in response to those flaws, and it is enabled by abductive reasoning [3]. The approach to modeling taken by Bryant et al. [2] is to frame binary RE as a sensemaking task, in which the engineer develops a mental model or hypothesis about the situation and its constituent elements, and adapts that model through iterations of goal-directed information seeking. Through a process of cognitive task analysis and verbal protocol analysis, they identified nine sensemaking functions in binary RE and developed a state machine representing the process, represented in Figure 2.2. Within their model, Bryant et al. included both procedural and declarative knowledge. Procedural knowledge consists of stored patterns of interaction. They categorized the declarative, or factual, knowledge into twelve subdomains covering the fundamental training and experience needed by reverse engineers (programming, debugging, program loading and execution, instruction sets, etc.). Additional declarative knowledge comes from abstract and concrete causal relationships within the data, e.g., constraints on prior knowledge or predicates applied to constants. They explored codifying declarative and procedural knowledge as an implied state machine via production rules implemented in the ACT-R [31] cognitive architecture.

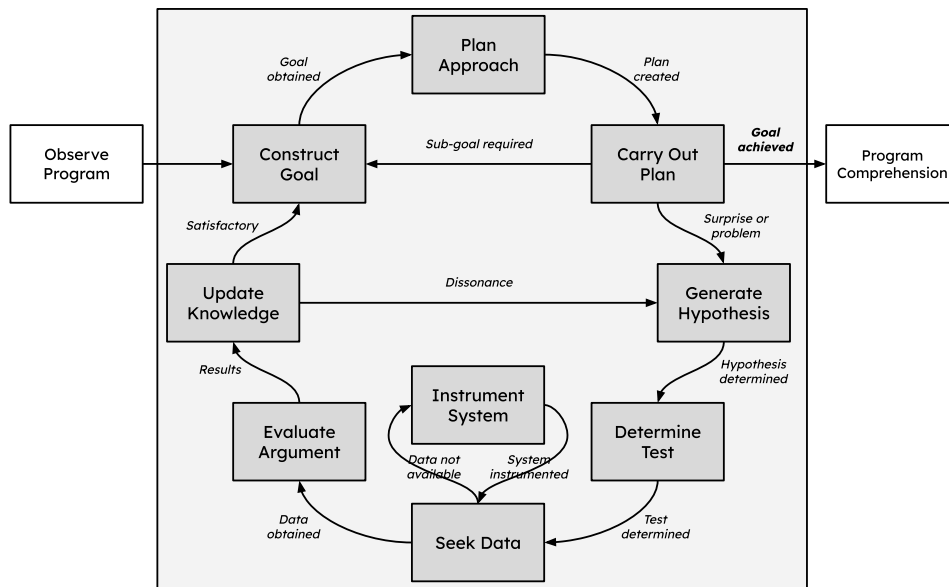


Figure 2.2: RE process based on sensemaking; adapted from Bryant et al. [2] and reframed as the central process in program comprehension / reverse engineering

To facilitate and encourage collaboration through communication of information in analysts' mental models, Tennor [32] surveyed and analyzed cognitive aspects of binary RE and

proposed that the binary reverse engineering community build vocabularies and ontologies. In this vein, Sisco et al. [33] sought to mathematically formalize Bryant et al.’s concept of binary RE as a sensemaking task and developed a supporting ontology. They asserted that reverse engineers analyze programs using four foundational patterns: navigation (searching for items, e.g., beacons ¹); translation (determining how the code would be implemented in a higher-level language); experimentation (deducing how program values change over different flows); and elaboration (identifying and explaining the major components and properties of the program). Reverse engineers use these patterns to build knowledge as interrelated mental objects with various constraints on their relationships. They proposed an ontology for representing this knowledge composed of *ologs*: category-based types, aspects, and facts; this representation allows commutation and a higher level of expressiveness than is supported by common alternatives such as Web Ontology Language (OWL) and Resource Description Framework (RDF). The team generated ologs for fundamental assembly instructions, program data, control flow, and operating system events, and used those ologs to formalize information flow in the experimentation pattern. The team theorized that this work provides the basis for a computational framework to model cognitive tasks in binary RE, and that model can further inform the actions of an analyst or agent encoded in ACT-R.

In a push toward building automated agents to assist with binary RE, Dudenhofer [3] proposed the Cognitive Understanding of Reverse Engineering (CURE) model to capture sensemaking steps in binary RE. The model, shown in Figure 2.3, is founded on the iterative cycle of abductive reasoning and experimentation: form a hypothesis; explore to find information to prove or disprove a hypothesis; recognize cues in the code or related artifacts, e.g., beacons, and store them in working memory; use that information to refine the hypothesis or form the next hypothesis; and repeat the cycle. The model inspired an application, the CURE Assistant, which works with an existing binary reverse engineering framework to identify code snippets

¹The *beacon* is one important concept in the cognitive process identified by Brooks, which is a “set of features that typically indicate the occurrence of certain structures or operations within the code.” [27]. An example of a beacon is a block of code that interchanges values within an array inside of a loop, which strongly indicates a sorting function. Programmers seek beacons to confirm their hypotheses in the abductive reasoning process.

matching those in an extensible catalog of recipes and present possible program behaviors to the analyst.

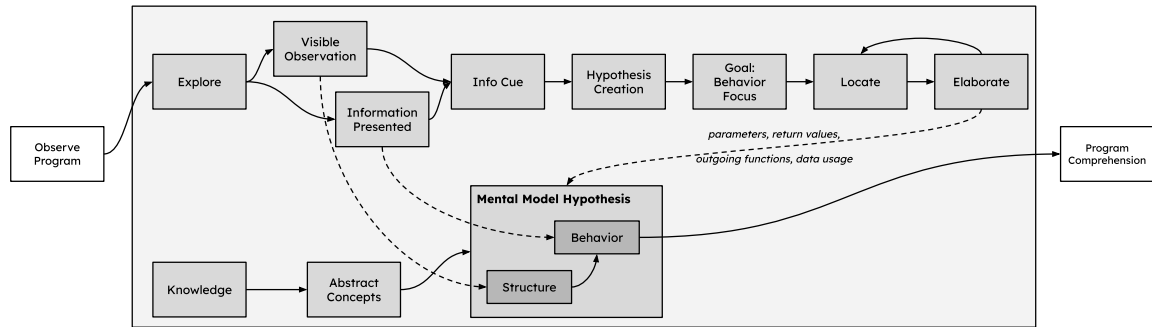


Figure 2.3: RE process based on abductive iteration; adapted from Dudenhofer [3] and re-framed as the central process in program comprehension / reverse engineering

Votipka et al. [4] presented a three-phase process model for binary RE: (1) Examining and executing the full program for an overview, then choosing focus areas; (2) Reviewing specific program slices chosen in the first phase, scanning for beacons and data and control flows, and generating specific hypotheses; and (3) Inspecting lines of assembly code and traces to test the hypotheses, as shown in Figure 2.4. They also identified several categories of beacons for use across the phases: APIs and strings across all phases; UI elements in the first phase; and constants, variable names, control flow structures, compiler optimizations, function prototypes, and program flow in the second and third phases. In developing this process, they identified one significant similarity between binary RE and source-code-based RE: mental simulation of code execution; and two primary differences between binary RE and source-code-based RE: binary RE involves more overview while source-code-based RE is more focused and pinpointed, and binary RE uses a more diverse set of beacons, common recognizable schema or patterns. Their work culminated in five guidelines for designing RE tools: (1) Match interaction with analysis phases; (2) Present input and output in the context of code; (3) Allow data transfer between static and dynamic contexts; (4) Allow selection of analysis methods; and (5) Support readability improvements.

Nyre-Yu et al. [34] conducted a task analysis of static binary RE. They found evidence reinforcing previous findings that binary RE and source-code-based RE share similarities in cognitive processes: identifying goals and plans, creating hypotheses, and exploring to gather

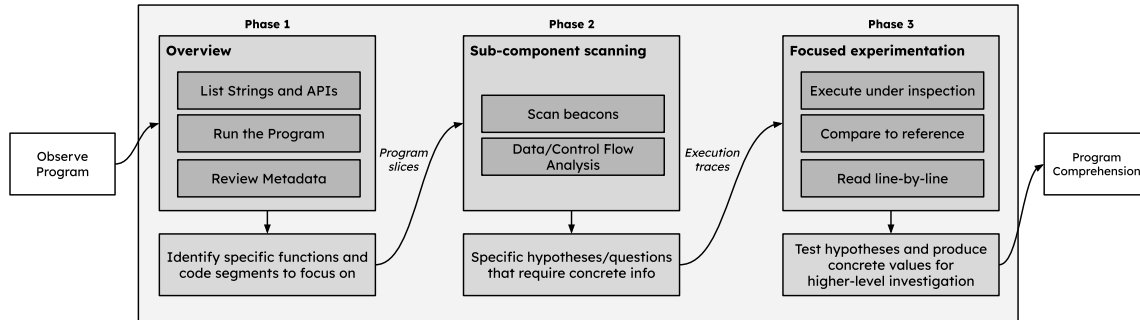


Figure 2.4: RE process in three phases, adapted from Votipka et al. [4] and reframed as the central process in program comprehension / reverse engineering

information. They asserted that analysts very commonly ask “where is this used in the code?”, “where is the method being called?”, “how can I get calling information?”, “where does this information/data go?”, “where is the data coming from?” and “what is the context of this vulnerability/code?” [34]. They also observed the revisiting of past states reported by Votipka et al. [4], however, they did not find common patterns across participants. The team also identified repetitive actions that are targets for automation: accessing library function documentation and accessing the task definition.

2.3.2 Differences between Novices and Experts in Performing binary RE

Besides the mental models associated with binary RE, it is important to understand differences between novices and experts. While both classes of analysts use many of the same elements of the mental model, the approaches they take are different.

Looking toward source-code-based RE again as a close analog of binary RE, Vessey [35] studied experts and novices in a code comprehension and debugging task, observing that while both experts and novices employ breadth-first approaches, experts apply a system view that novices do not; novices also employ depth-first approaches while experts do not. Additionally, experts perform chunking effectively to proceed steadily through a program, while novices perform chunking much less effectively, resulting in jumping around within a program to understand it. Storey’s survey [36] found that experts use more external memory aids [37], and that novices focus mainly on objects while experts also consider functional relationships and algorithms. Siegmund et al. [38] asserted there are two basic models of RE: top-down and

bottom-up. Rather than distinguishing between “expert” and “novice,” they focused on level of domain knowledge. Analysts who have domain knowledge they can apply to the program use the top-down process and use beacons to form hypotheses. Otherwise, without domain knowledge, they use the bottom-up process to analyze the program line-by-line.

Cowley [39] conducted a job analysis to identify performance predictors for binary RE practitioners. In addition to situational (team and organizational) predictors, Cowley identified individual predictors for novices and experts. From those predictors, the author determined five milestones along the progression from novice to expert. Milestones to the intermediate level include (1) proficiency in relevant tools; (2) significant reduction in assistance needed to complete tasks; and (3) parity with experts in identification and employment of binary RE strategies. The final two milestones complete the transition to a true expert, including (4) organizational recognition through promotion; and (5) establishing a track record of solving binary RE problems without assistance.

Mantovani et al. [12] studied nearly 300 hours of activity performed by 72 novice and expert reverse engineers performing a static analysis task. They observed that most often, novices move forward from `main()` and jump around between code blocks, visiting some more than once, while experts move both forward and backward from `main()` and move more linearly through code blocks. Additionally, experts are quicker to identify what they can ignore.

2.3.3 Section Summary

In this section, we surveyed research in support of answering RQ1, “What are common and significant characteristics of mental/cognitive models employed by binary RE practitioners?” We examined a progression of research contributions in this area. While researchers found that the process of binary RE is uniquely crafted by each individual, especially at the expert level, we identified several elements cutting across the research. Analysts commonly follow a process of abductive iteration: setting goals, forming hypotheses, testing, and revising based on findings, iteratively making sense of low-level code. External memory aids are critical, while

diverse cognitive cues (beacons) guide the analyst. Analysts typically translate low-level code to higher-level representations and filter out what seems irrelevant to the current goal. There are clear differences in how novices and experts analyze a binary program. We revisit these elements in Section 2.6, Synthesis, where they are presented in a granular and attributed format in Column A of Table 2.1.

Continuing our effort to understand the circumstances of the problems of binary RE, we next look at findings in the literature about underlying cognitive theory with the aim of augmenting or improving the cognitive activities of binary RE.

2.4 Underlying Cognitive Theory and Applications to Binary RE

Now that we have addressed the characteristics of mental/cognitive models behind binary RE, we seek to understand the circumstances and causes of difficulties practitioners encounter (CSE step 2). In this section we answer RQ2, “What are underlying elements of cognitive theory bridging sensemaking in binary RE and immersive sensemaking?” We briefly review the underlying theory behind external and embodied cognition along with cognitive load, and then consider methods that may reduce cognitive load or otherwise optimize the cognitive processes of binary RE practitioners.

2.4.1 External and Embodied Cognition

Cognition involves many specific and interdependent processes: Eysenck and Brysbaert [40] identified attention; perception; memory; learning; reading, speaking, and listening; and problem-solving, planning, reasoning, and decision-making; each with their own design implications. Norman described cognition as multimodal—that there are many different types of thinking—and specifically addressed experiential cognition and reflective cognition [41]. Experiential cognition is when a person is experiencing and responding to the environment without the need for significant mental effort (which also happens with extensive expertise), while reflective

cognition happens when a person is putting substantial thought into considering and making decisions or forming new ideas. Norman further discussed both forms in relation to effective tools for enhancing one or the other, such that a tool designed to aid reflective thought is inappropriate for experiential cognition and vice versa [41]. Cognition also occurs in varied places, times, and situations, and there is movement in the field toward understanding cognition *in situ*—rather than limiting the scope of reasoning about cognition to what occurs in the mind, considering how the environment can improve and affect cognition [24].

External cognition, per Scaife and Rogers [42], concerns the cognitive interaction with external manifestations of knowledge in the environment, e.g., images, video, virtual reality, etc. The primary cognitive benefits of external cognitive activities include using external knowledge representations (e.g., notes and reminders) to reduce memory load; using computational tools (e.g., calculators) to make tasks easier; and annotating and reordering or restructuring external representations of knowledge (e.g., checking off to-do lists or arranging desktop icons) [24]. In considering the impact of the environment on cognition, Gibson coined the term *affordance* as a “specific combination of the properties of its substance and its surfaces” perceived in relation to the viewer [43]. Norman further elaborated that an affordance is most importantly constrained by properties that indicate how something could be used: “When affordances are taken advantage of, the user knows what to do just by looking: no picture, label, or instruction needed.” [44]. Affordances leverage internal cognition to enable the phenomena of external cognition and embodied cognition.

Embodied cognition [45], closely linked to dynamical systems theory, proposes that cognition happens in real time, with the brain simultaneously receiving input from, processing, and interacting within and with the nervous system, physical body, and external environment. In this theory, cognition is deeply impacted by sensorimotor interaction with the environment and is profoundly grounded in the ability to act [46]. Shapiro and Spaulding’s *Embodiment Thesis* states: “Many features of cognition are embodied in that they are deeply dependent upon characteristics of the physical body of an agent, such that the agent’s beyond-the-brain body plays a significant causal role, or a physically constitutive role, in that agent’s cognitive processing.” [47]

In line with the concept of the joint cognitive system of human and machine described earlier, Kirsh [48] made the case that tools are absorbed into the body schema and change how we think, that we think with not just our minds but also our bodies (enabled by kinesthetic perception), and that we even think with tools. Similarly, Hornecker et al. [49] presented the position that our sensorimotor interactions with, and manipulation of, the world develop our capacity for abstract thought, starting with simple concepts such as in/out, over/under, up/down — that body and mind are inseparable even in the domain of abstract thought. Ale et al. [50] suggested that embodied memory is a potentially rich research area, in which physical objects or locations serve as memory palaces; and that whole-body stimuli can expedite storage and retrieval of memory (e.g., using physical motion or aroma triggers to encode and store memories).

2.4.2 Cognitive Load Theory

Cognitive Load Theory (CLT), per Sweller and Chandler [51], assumes that humans take in and process information through two main channels—through hearing and through visualizing. Further, only a finite amount of processing can occur in each channel at any point in time. The magnitude of the cognitive load depends not only on sheer number, but on how much interactivity occurs between elements. The ability to process this information is limited by working memory, but can be improved through employing a suitable schema with which to leverage long-term memory and reduce cognitive load. Cognitive load theory currently incorporates three categories: *intrinsic* load inherent in the cognitive task at hand; *extraneous* load caused by how information is presented; and *germane* load. Germane load represents the ability of our mind to connect what we are learning with long term memory, and linked to intrinsic, is the demand on our cognition of using or focusing working memory for intrinsic learning [52]. Paas et al. [53] asserted that cognitive performance will degrade at either end of the load spectrum (underloading or overloading).

Hollender et al. [54] surveyed 65 papers on cognitive load related to human-computer interaction. They cataloged methods to leverage phenomena to reduce extraneous cognitive load

as follows. The worked example effect is from learning from studying solved sample problems. The split-attention effect is from presenting information from multiple visual sources in an integrated way to reduce load required to perform mental integration. The modality effect is from presenting multiple information sources through different modalities (e.g., visual and aural) to allow the inputs in each modality to be processed simultaneously. The redundancy effect is from reducing the level of redundant information presented in different modalities/sources, thereby reducing the load of reconciling the underlying concepts across the inputs from those modalities. They also reviewed methods to increase germane cognitive load and foster schemata development (most applicable in educational settings, and a way to increase the capacity of working memory): specifically introducing a variety of tasks; linking concrete information to abstract concepts; and self-explanation. Also, in an educational context, they reviewed methods to adjust intrinsic cognitive load via adjusting the sizes and quantity of information chunks presented over time.

2.4.3 Applications to Program Comprehension

Next, we will review contributions from studies applying underlying cognitive theory to RE and closely-related tasks. Helgesson and Runeson [55] studied cognitive load drivers in software engineering and proposed a set of perspectives with which to reason about these drivers. The *Task* perspective accounts for CLT's intrinsic load, which results from the inherent cognitive intensity of software engineering. The *Environment* perspective accounts for CLT's germane load, which results from constructing mental schemata for processes and tools, plus additional load from re-learning new processes and tools that are meant to replace old ones, but can set up competing mental schemata. The remaining perspectives comprise CLT's extraneous load: *Structural* (e.g., technical debt); *Information* (e.g., poor/missing code documentation); *Tool* (e.g., friction from unintuitive, cumbersome, or unreliable tools), *Communication* (e.g., lack of communications amongst the development team), *Interruption* (the cognitive cost due to resumption lag), and *Temporal* (e.g., tracing a component's change history in version control).

In observing how programmers comprehend code, Siegmund et al. [38] took a novel approach by employing functional magnetic resonance imaging (fMRI) to find the brain regions activated during source-code-based RE. In a controlled study of 17 computer science undergraduates, designed to elicit bottom-up source code comprehension and minimize extrinsic cognitive load, they observed activation of five Brodmann areas (of 52 Brodmann areas associated with cognitive process) [56] using fMRI. Those five areas are associated with division of attention, silent word reading, verbal/numerical working memory, and problem solving. Based on further analysis, Siegmund et al. theorized that bottom-up RE uses two areas for keeping values in mind, another area for analyzing words and symbols, and the remaining two areas for integrating statements and chunks.

Smits [14] performed a study of reducing cognitive load in reverse engineering. Under the assumption that the complexity and volume of outputs from RE tools induces cognitive overload, Smits implemented two primary techniques for managing cognitive overload: (1) *Information Filtering* to remove extraneous information so as to reduce extraneous load; and (2) *Information Organization* to organize data in a manner familiar to the user to reduce germane load. The two techniques would therefore increase the capacity available for intrinsic load. The implementation focused on improving the common Control Flow Graph (CFG) visualization with the concept of the *Proximity View*: Simplify the view by removing most instructions, variables, and constants; keep only variables and constants that are arguments for function call nodes; and insert empty nodes to maintain the graph structure. A user study of 41 participants comparing this view to a traditional view demonstrated that subjects in the Proximity View group had statistically significant better performance in challenges solved, but took longer to solve them.

2.4.4 Section Summary

In this section, prompted by RQ2, “What are underlying elements of cognitive theory bridging sensemaking in binary RE and immersive sensemaking,” we reviewed relevant cognitive theory and its applications to binary RE and found several common and salient elements.

One key idea is external cognition, offloading information to external representations to reduce mental effort. More directly relevant to VR, the concept of embodied cognition proposes that we shape and extend our cognitive processes through interactions with tools and spaces. Further, spatial arrangements can enhance memory per the concept of embodied memory. Finally, cognitive load theory suggests mechanisms to optimize mental effort (especially to reduce extraneous load) such as multi-modal information presentation, reducing redundancy of information presented, and adopting human-centered design. We revisit these elements in Section 2.6, Synthesis, where they are presented in a granular and attributed format in Column B of Table 2.1.

Now that we have reviewed elements of cognitive theory that underlie the cognitive processes of binary RE, we will examine how immersive technologies can improve those cognitive processes.

2.5 Cognitive Augmentation of Sensemaking Using Visualization and Immersion

The third step of CSE directs that we should pursue practical solutions to the problem areas that we have identified. In pursuit of such solutions, we present prior work in applying immersive technologies to improve cognition and performance in analytic tasks in general, with a few examples specifically for program comprehension and understanding. This section answers RQ3, “How have the affordances of immersive virtual reality been employed to improve cognition and sensemaking in analytic tasks?”

2.5.1 Leveraging Affordances of Immersive Technologies for Analytic Tasks

As described in Section 2.4, affordances are perceivable aspects of the environment that foster interaction and enable external and embodied cognition. Some common affordances of virtual, augmented, and mixed reality (VR, AR, MR) include immersive visual and aural displays, spatial interaction / motion tracking, gesture and voice recognition, haptic feedback,

plus additional emerging modalities. In this dissertation, “affordance in VR” refers to an environmental feature available in the VR experience, including features that are themselves intrinsically 2D, such as code listing windows, that can be created, sized, oriented, and placed arbitrarily in the 3D space by the user to carry out a task; this inclusion builds a foundation for evaluating the effects of external and embodied cognition and is the subject of active research in other use cases [57]. In this subsection, we look at prior work in using the affordances of immersive technologies to improve cognitive processes used by analytic and sensemaking tasks.

One key feature of an immersive environment is an abundance of space. Although the work of Andrews et al. [58] focused on the use of large two-dimensional displays, they proposed several benefits that working in a large spatial environment brings to the sensemaking process. They completed an observational study of participants performing well-known data analytic tasks using their desktop computing environment, which included a 32-megapixel display of approximately 100 inches by 31 inches. Generally, participants used the space to arrange and organize documents and applications in ways that reflected their relationships. The study showed evidence of a number of avenues to exploit for improving cognition by using a large spatial environment including persistence, context, physical navigation, presence of detail, memory refresh, situational awareness, and spatial semantics. These concepts are listed in full in the section summary, citing the authors, and will be further explored in Section 2.6. The findings by Andrews et al. endorse concepts we covered in Section 2.4 regarding external cognition and memory (augmenting working memory and structuring external knowledge representations) and embodied cognition (exploiting and extending the body schema).

Follow-up work in this area considered the effect of available space on sensemaking, but in immersive 3D. Lisle et al. performed observational studies [59], [57] of users examining large sets of historical documents. Participants donned a VR headset to select, read, and organize 2D documents in a virtual room in support of performing analysis tasks as directed. The researchers identified three main organizational structures used by participants (semicircular, environmental, and planar). Participants demonstrated higher performance when they used labels and highlights, examples of external cognition, during task execution. Further, Derksen et

al. [60] conducted a user study comparing task performance (retrieving scatter plots) in virtual environments with varying levels of environmental features. They found that the time and effort users put into arranging the plots had a greater positive impact on performance than the richness of the virtual environment, suggesting that allowing users to organize their space supports spatial memory in immersive analytic tasks.

In their hybrid survey/position paper, Moloney et al. [61] consider the question of whether VR technology provides affordances for analyzing big data. Their survey covers 47 selected publications in the field of visual data mining, which employs visual cognition to augment algorithmic analysis by differentiating variables by mapping them to distinct graphic attributes to harness human visual perception and creativity. Starting with the Computer Aided Virtual Environment (CAVE) in 1992, they provide a timeline of advances through ‘VR 2.0’ and IA, and present future research challenges proposed by research teams. The authors perform an analysis of affordance theory in immersive VR as it exists in the literature and present a position on the shift from allocentric (attention focused externally) visual analytics—traditional visual analytics [62], which is performed using a 2D screen, even for 3D visualizations—to egocentric (attention focused internally) spatial coding and the affordance of VR. The key principles they identified include tuning the environment to human perception; using mimetic/-naturalistic references; coordinating multiple modes of interaction; and aligning data selection with naturally-occurring distribution patterns. These principles are listed in full in the section summary, citing the authors, and will be further explored in Section 2.6.

Gračanin [63] makes the case that VR and MR technology can provide significant advantages in forming insights about complex data sets by leveraging the theory of affordances to increase embodied resources. The author proposes a framework for creating stimuli in an MR environment based on data from sensors and the user’s interactions with the system, which can reveal the most effective immersive and embodied stimuli through an iterative feedback process. Gračanin concludes that the immersion provided by VR may not be sufficient to fully leverage embodied cognition, and that the physical/real-world affordances provided by MR resolve that gap. Billingham et al. [64] similarly have developed a VR system incorporating a feedback loop using sensor data to adapt the interaction. Their work has demonstrated that

electroencephalogram (EEG) measurements can assess cognitive load in VR training tasks, and the simulations can be adapted based on those measures to benefit training transfer effects. Ahmadi et al. [65], in related research, further refined the findings to pinpoint that power spectral density measurements of the EEG alpha band may reliably indicate cognitive load in moderate VR tasks.

Batch et al. [66] conducted an experiment to determine the effects of using IA for tasks in economic analysis. They used an iterative, human-centered approach with subject matter experts to extend the ImAxes tool [67], which provides embodied data axes in a VR environment. They employed a VR headset instead of AR due to better resolution and field of view. Some of their findings were surprises—e.g., fatigue was not a significant factor; no clear concerns about legibility; and participants with little gaming/VR experience used the tool effectively. Participants created egocentric presentation layouts, but did not appear to use the physical space to group similar views near each other, simply using the closest free space. Although this observation contradicted their hypothesis, the authors assert that what they did observe is consistent with the ‘sensemaking loop’ described by Pirolli et al. [68]: in the early bottom-up search for information when executing a task, retained information is left somewhat unstructured except to store it in a “shoebox” for later processing. These findings of Batch et al. corroborate those of Andrews et al. [58] when employing large two-dimensional displays for sensemaking tasks. Ultimately, Batch et al. [66] concluded that participants still did not fully use the three-dimensional space available in the environment beyond their immediate proximity, and speculate causes may include the small size of the physical space in which the study was conducted; ability to perform the techniques or gestures required to move the visualizations; and lack of automatic layout, indicating a need for constraints and organization frameworks.

Prather et al. [69] surveyed 104 papers on cognitive augmentation using immersive technologies, specifically work employing biosensor-based measures of user cognitive capabilities in immersive and semi-immersive environments. The authors performed this survey to support their aim of designing and developing cross-reality (XR) systems where task parameters are adapted to optimize the user’s cognitive load, particular in ‘Industry 4.0’ use cases where machines carry out repetitive and increasingly complex tasks. These systems would perform

as intelligent cognitive assistants to enhance human capabilities. Their survey found that existing research is predominantly in the health and rehabilitation domain rather than industrial engineering. Many efforts target affective/emotional wellness with an emphasis on short-term therapeutic tools rather than an enduring assistive system. Over one third of papers described work in adapting the user experience based on physiological data (predominantly electroencephalography (EEG)), with a small but significant number of them applying Artificial Intelligence (AI) and Machine Learning (ML) to that data. The authors included no papers that directly addressed CLT, but several addressed mitigation of mental exertion.

Some recent experimental findings in research on immersive technology and cognitive load pinpoint where the technology is most and least effective. Frederiksen et al. [70] performed a study of cognitive load in surgical training using immersive and non-immersive VR. The immersive VR method increased cognitive load significantly more than non-immersive VR in both nonstressor and stressor phases, which was attributed to extraneous cognitive load due to the high level of interaction with immersive VR elements. De Melo [71] et al. sought to reduce cognitive load with a virtual embodied AI-based assistant that used a human-appearing avatar in the virtual world to provide guidance to participants in completing tasks. In a controlled experiment comparing the embodied assistant to voice-only assistant and no assistant, embodied assistants led to lower cognitive load than voice-only assistants and both were lower than no assistant. Albus et al. [72] studied the effects of the signaling principle in VR learning environments. Per Mayer [73], using signals to direct a user to pay attention to specific information can create deeper understanding; this is the signaling principle. Compare this concept to beacons described earlier. In their study of a learning exercise with and without annotations (signals), Albus et al. found that the signals increased germane cognitive load, but did not reduce extraneous cognitive load, and did not result in significantly better deep understanding. Chen et al. [74] performed a controlled experiment measuring engineering creativity in a secondary education setting, with and without VR. They found that while VR improved cognition, motivation, and the novelty and usefulness of the designs, it did not improve creative thinking. Additionally, VR improved extraneous and germane cognitive load, but had no effect on intrinsic load.

The field of IA emerged only recently [16], [17] with an aim to immerse users in their data to improve sensemaking and formation of insights for real world data analytic tasks. Most recently, Ens et al. [18] have proposed ‘Immersive Analytics 2.0.’ They recognize that evidence found in prior work in IA points to benefits resulting from several areas: “use of unlimited space around the users, spatial memory, proprioception [awareness of body motion], 3D spatial interaction, tangible interaction [props and haptics], collaboration, kinesthesia [body motion], [and] engagement.” They call for work to more deeply understand the motivations for spatial and embodied interaction, and to clearly compare immersive versus non-immersive experiences in analytics.

2.5.2 Visualization Applied to Software Reverse Engineering

Employing visualization can impact the effectiveness of analysts in binary RE and how they exercise their cognitive models. This section covers prior work in visualization to support RE, including that for binary RE and applicable work in source-level RE.

The bulk of prior work in this application area is aimed at the 2D desktop interaction metaphor, and we provide a few notable examples. Early work by Waguespack [75] implemented visualizations to aid novice Pascal programmers, for example, representing data types as different shapes, and using a variety of visual representations to differentiate type declarations, variable instances, and literal values. Structures are represented as containers of constituent components. This work demonstrated the utility of *chunking*: collecting lower-level details in a single higher-level abstraction that can help understanding of a large problem, then providing the ability to decompose as necessary for detailed analysis. Conti et al. [76] implemented a 2D application to visualize binary and data files as bitmaps; e.g., representing one byte per pixel with shading based on the byte’s value or presence in an address range. Rendering a binary file in that way, compared to viewing it in a hex editor, gives a view of the entire file at a glance at the expense of low-level details. The method complements the text-based viewer to quickly identify major segments in the file, recurring or unusual patterns, and so on.

Greggio et al. [77] demonstrated two 2D/pseudo-3D visualization methods of behavior of suspected malware: timeline plots and icons arranged in a spiral. Both methods display operating system actions taken by the code on objects (e.g., read file; terminate process; etc.) over a period of time in a single, compact view that complements traditional logs. The visualizations were particularly useful in identifying where two sample programs acted similarly, indicating shared code across different programs, or different revisions of the same program.

Considering a narrower application area, Wagner et. al. [78] surveyed a pool of 220 papers related to code visualization and identified 25 papers specifically about malware visualization systems. From those papers, they identified nine *data providers* behind those visualization systems. The data providers collect information about the suspected malware and provide it to the visualization systems—automated and manual applications that execute static and dynamic analysis techniques to collect data useful in profiling and classifying the malware. The visualization systems were binned into three broad groups: individual malware analysis, malware comparison, and malware summarization. Additionally, they were categorized based on well-established taxonomies from the visualization community: the type of provided data, visualization techniques used, mapping and representation space, temporal aspects, interactivity, and goal/action. They identified challenges in bridging between the three broad groups, integrating disparate data sources, characterizing and abstracting problems, improving expert interaction, and integrating analytical methods with the visualizations. These systems are still aimed at the 2D desktop interaction metaphor—while some of the systems analyzed employed 3D visualizations, none were reported in this survey to employ virtual reality or immersive (embodied/egocentric) techniques.

Elliott et al. [79] explored the use of VR in software engineering to address problems in navigating and comprehending code. Their work builds upon prior research in how developers use the affordance of *spatial memory* in traditional 2D development environments, such as using scrollbar and tab positions as cues [80], or using an “infinitely”-scrollable document canvas [81], [82]. This work extends that concept to the affordances of VR applied to software development: spatial cognition, cues, and presence; manipulation and motion to improve

perception and retention; and immediate feedback on the state of the system. With these affordances implemented in their RIFTSKETCH (live development) and IMMERSION (code review) tools, the authors provide a proof-of-concept and a vision of VR-based development in the future, though no formal user studies were conducted.

In a user study conducted by Dominic et al. [83], 26 graduate students were tested on comprehending simple Java programs of the type that one may find in first-year programming course homework assignments. They compared the traditional desktop experience with a VR configuration entitled “VirtualDesk” using a headset and tracked keyboard and mouse that were mapped 1:1 to the real world. The study did not implement specific affordances of VR as proposed by Elliott et al. [79], but instead compared the performance of RE using common 2D tools on a traditional desktop environment against using the same tools in VR. Their results show 75% of programs were comprehended correctly in the traditional desktop experience compared to 65% in VR. Additionally, their results from conducting the NASA Task Load Index (TLX) survey [84] showed significantly more task load—the demand or difficulty in performing a task—in VR. Finally, results of a survey of self-reported concentration and productivity showed that users in VR had lower levels of concentration and no significant difference in perceived productivity.

One rich facet of human-computer interface theory is the use of metaphors to influence the design of affordances. Lakoff [85] curated an extensive list of metaphors encountered in linguistics. Metaphors leverage common experiences amongst most users to facilitate understanding of new concepts. Many of these metaphors can apply to perception and cognition in interactive applications, e.g., “seeing is touching,” “the visual field is a container,” “theories are constructed objects,” etc. Averbukh et al. [86] surveyed applications of VR for (high level) program visualization and visual programming, and in particular, the metaphors employed in those applications. They reviewed the city, molecule, and heliocentric cosmic metaphors, asserting they share important qualities: “unlimited context, organization of inner structure, naturalness, and resistance to scaling,” and that these natural metaphors simplify spatial orientation and navigation in the VR world.

The city metaphor recurs in many VR-based visualization efforts. One early instance was by Fittkau [87] et al., who implemented a VR experience to aid the RE process that uses the metaphor of a program as a city block: the buildings are classes and packages, and the execution trace is represented as straight-line “footpaths” between the buildings. Participants experienced the tool ExplorViz via immersive VR, using gestures to translate, rotate, zoom, and select. The experience was intended to provide analysts a novel tool while employing familiar metaphors. These participants rated the experience of answering basic comprehension questions with this tool as suitable for performing RE, and as an alternative, albeit needing adaptations, to a classic experience.

In another exploration using the city metaphor, Oberhauser and Lecon [88] employed immersive VR to aid RE by providing participants the ability to fly through a 3D representation of code in their tool Gamified Virtual Reality FlyThruCode (GVR-FTC). Two metaphors, “universe” and “terrestrial,” related the code components to familiar concepts, where packages, classes, and dependencies were represented as solar systems, planets, and light beams, or cities, buildings, and pipes respectively, and the scale of objects represented various metrics such as number of class methods. The team evaluated the tool using two games that motivated players to comprehend the dependency structure and modularization of a code project compared to a common text editor. Although the subject pool was too small to demonstrate statistical significance, the results did show higher comprehension using VR.

In a larger-scale experiment, Romano et al. [89] compared the relative effectiveness of three tools on the task of source code RE. The baseline was a traditional integrated development environment (IDE) with extensions for code metrics and smells, which was compared to a city-metaphor-based virtual reality environment in both immersive (Code2City_{VR}) and non-immersive (Code2City) forms. This implementation of the city metaphor [90] creates a building (parallelepiped) for each class, where class properties are reflected in the size and color of the building. In their study [89], 42 participants solved RE tasks based on two large open-source Java projects. Both VR-based tools resulted in significantly better correctness in the completion of RE tasks than the IDE. Additionally, the time to complete the tasks was significantly shortened in the immersive environment than both the non-immersive VR and the IDE.

More recently, Hoff et al. [91] performed a similar experiment for source code RE comparing their approach, Immersive Software Archaeology, with another VR method and an IDE. Their approach is focused on providing an overview, with multiple levels of abstraction, of a software system’s architecture. The higher levels of abstraction (architectural) were represented by solar system/planets/continents, and the lower levels (design) were represented by cities/building/floors. Their study of 54 participants demonstrated that their solution provided similar or better performance in tasks exercising accessing information and finding horizontal and vertical relationships in the system’s architecture. Further, this team implemented collaborative note-taking features in their immersive code analysis system [92], adding shareable drawing, audio recording, and screenshot capabilities. An initial case study showed that participants found the system beneficial and they produced correct results.

Finally, we mention a variation of the city metaphor. Weninger et al. [93] introduced the concept of *Memory Cities* to visualize how an application uses heap memory over time, rather than using it to visualize code. Objects on the heap are grouped and represented as buildings in a 3D visualization. Attributes such as color, opacity, area, and height represent various metrics of the heap, and the buildings evolve as the program executes. The authors describe how their tool helps users identify memory leaks in two use cases and plan user studies in the future.

As mentioned at the start of the section, we are careful to distinguish research contributions that focus on source-level RE from those that focus on binary RE. We include them because, even though we will not have higher-level structures, abstractions, and internal documentation for a binary program, and even though any higher-level abstractions generated by tools will have significant uncertainty, we believe many of the findings presented in this subsection can inform immersive VR for binary RE.

2.5.3 Section Summary

In approaching RQ3, “How have the affordances of immersive virtual reality been employed to improve cognition and sensemaking in analytic tasks,” we reviewed relevant research contributions in the fields of immersive technologies as well as visualizations and metaphors

employed in RE. Spatial organization introduces a semantic layer supporting memory and understanding through the arrangement of items in 3D space. This organization, through persistence and context, helps users re-enter prior states of mind, enhancing recall and continuity. Embodiment extends cognitive processes into the environment through physical navigation, gestural interaction, and embodied assistants. Environment design enhances situational awareness by facilitating scanning the space and glances that refresh memory. Representations tuned to human perceptual strengths, such as cross-modal mapping, chunking, mimetic references, and using metaphors, further enhance understanding in the environment. We revisit these elements in Section 2.6, Synthesis, where they are presented in a granular and attributed format in Column C of Table 2.1.

Next, we will analyze our findings regarding VR affordances in the context of our prior findings in cognitive models of binary RE and the cognitive theory behind those models.

2.6 Synthesis

RQ4 question asks, “How can we use these findings to effect improvements in the practice of binary RE?” To answer this question, we examine the connections between the findings in the previous sections and identify the most salient themes in order to propose directions in the information and interaction design for this problem domain. This section continues CSE step 3, pursuing practical solutions.

2.6.1 Overview of Elements and Connecting Threads

Consider the findings of this survey partitioned into three groups matching the three previous sections: (A) characteristics of sensemaking / cognitive models of binary RE, (B) concepts of cognitive theory, and (C) VR affordances/tools/techniques for immersive sensemaking. Table 2.1 provides the full text of each element with attribution, by group. Conceptual *threads* (also called *strands* in similar work) tie an individual cognitive model element from group A to a cognitive phenomenon in group B demonstrated by the model element, and then to a VR technique in group C that, when implemented in an immersive binary RE tool, could improve

that phenomenon. Figure 2.5 provides an overview of these threads tying together the elements from each group (element labels are abbreviated). The authors determined these threads subjectively, linking elements between groups that the authors believe to be sufficiently directly related. We acknowledge that the threshold of “directness” is arbitrary and tuned to produce tight groups, and that strong arguments can be made to link the elements in other ways. We present the entire body of threads in the figure to illustrate just one take on the many logical avenues of further research this problem domain provides before we further narrow the set. Next, we will narrow our focus.

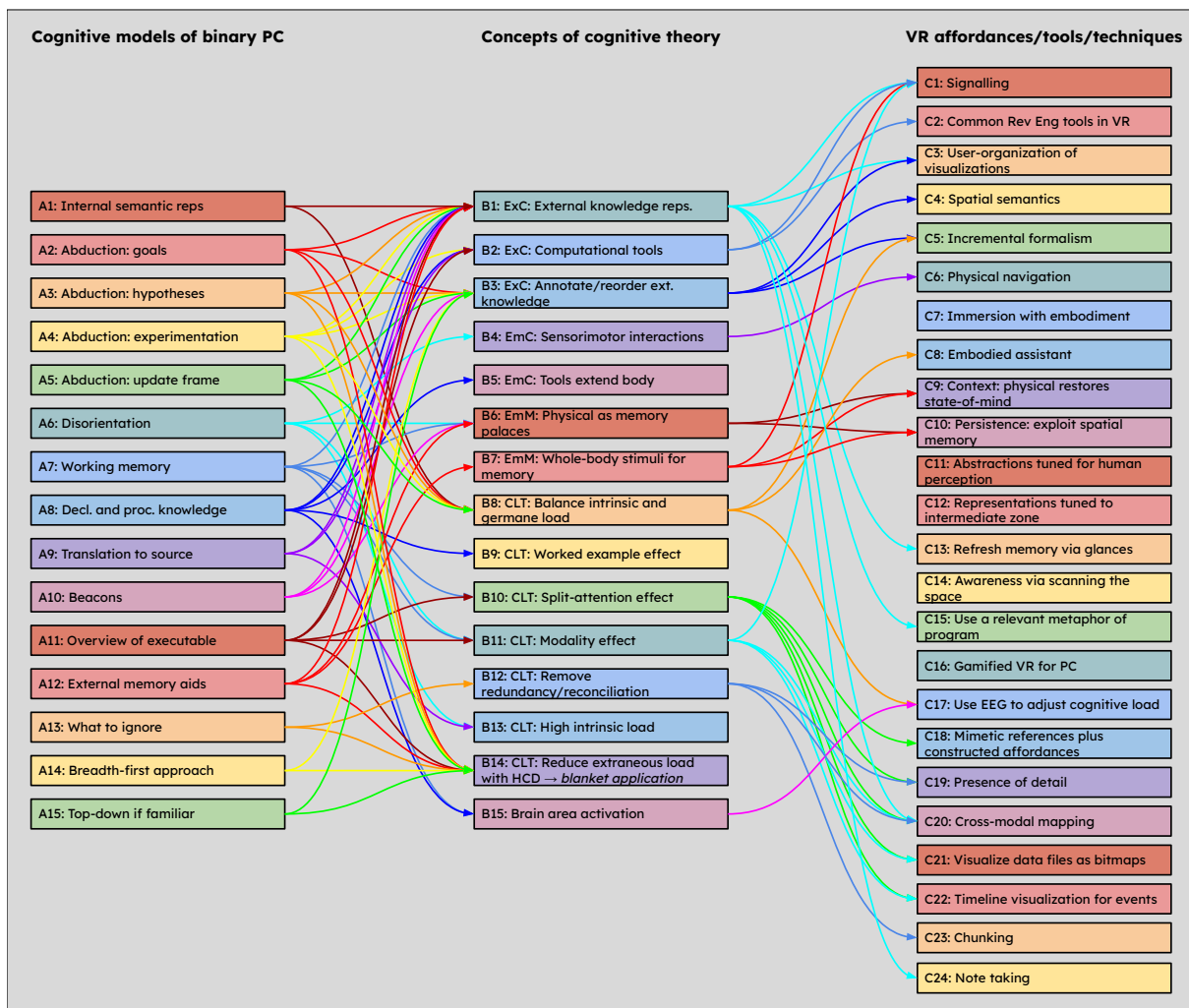


Figure 2.5: Threads connecting elements of (A) cognitive models of binary RE, (B) concepts of cognitive theory, and (C) VR affordances/tools/techniques

<u>Cognitive models of binary RE</u>	<u>Concepts of cognitive theory</u>	<u>VR affordances/tools/techniques</u>
A1. Uses multi-level internal semantic representation of the program [1] A2. Abductive iteration: Sets goals and follows plans [2], [34] A3. Abductive iteration: Forms increasingly complete hypotheses [27], [28], [3], [4] A4. Abductive iteration: Tests hypothesis through experimentation [2] [33], [3], [4] A5. Abductive iteration: Updates framing of the problem based on results [30], [2], [3], [4] A6. Creates disorientation following recursions and execution paths [29] A7. Taxes working memory [1], [29] A8. Requires declarative and procedural knowledge retrieval and generation [2] A9. Uses translation–determining how the code would be implemented in a higher-level language [33] A10. Uses beacons; beacons for binary RE are more diverse than for source-code-based RE [27], [3], [4] A11. Relies upon overview of the binary executable [4] A12. Uses external memory aids [37], [36] A13. Relies upon determining what to ignore [12] A14. Experts use a breadth-first approach with system thinking; novices, both depth- and breadth-first without system thinking [35] A15. Experts use a top-down approach if the program domain is familiar, otherwise, bottom-up [38]	B1. External Cognition: Use external knowledge representations to reduce memory load, e.g., notes and reminders [42], [24] B2. External Cognition: Use computational tools (e.g., calculators) to make tasks easier [42], [24] B3. External Cognition: Annotate and re-order or restructure external representations of knowledge [42], [24] B4. Embodied Cognition: Cognitive processing is influenced by the body and sensorimotor interactions [46], [49], [47] B5. Embodied Cognition: Tools extend the body schema [48] B6. Embodied Memory: Physical objects or locations serve as memory palaces [50] B7. Embodied Memory: Whole-body stimuli can expedite storage and retrieval of memory [50] B8. CLT: Balance immediate problem-solving (intrinsic load) and long-term schema development (germane load) [52] B9. CLT: Worked example effect of learning from studying solved sample problems [54] B10. CLT: Split-attention effect of presenting information from multiple sources in an integrated way to reduce load from mental integration [54] B11. CLT: Modality effect of presenting multiple information sources through different modalities (primarily visual and aural) to reduce the integration load [54] B12. CLT: Remove redundancy of information presented in different modalities/-sources to reduce the load of reconciling the underlying concepts [54] B13. CLT: Both underloading and overloading can degrade performance; underloading is unlikely due to the high intrinsic load of RE [53], [55] B14. CLT: Reduce extraneous load: Use human-centered design to reduce the wasted time and friction of poor tools and interactions used to solve a problem [55] B15. RE activates areas of the brain associated with working memory, written language, and integration [38]	C1. Signalling [72], [73] C2. Incorporate common reverse engineering tools C3. User-organization of visualizations in 3D space [66] C4. Spatial semantics: spatial organization provides added semantic layer [58], [57] C5. Incremental formalism: structure is emergent with understanding [58] C6. Physical navigation: enables efficient access to information through quick body movements [58], [57] C7. That immersion must be accompanied by embodiment [63] C8. Embodied assistant [71] C9. Context: physical location helps to restore state-of-mind [58], [57] C10. Persistence: exploits spatial (position and representation) memory to remember information [58], [57] C11. Use abstractions of environments to which human perception is attuned [61] C12. Use representations tuned to intermediate zone where human perception is most discerning [61] C13. Refresh: serendipitous glances refresh memory of information [58], [57] C14. Awareness: scanning the space quickly assesses overall status [58], [57] C15. Use a relevant metaphor to visualize programs (e.g., city block) [87], [88], [90], [86], [89], [91] C16. Gamified VR for RE [88] C17. EEG-based adjustment of cognitive load [64] C18. Mimetic references overlaid with constructed affordances [61] C19. Presence of detail: detailed information enables rapid access and synthesis based on rich content [58] C20. Cross-modal mapping [61] C21. Visualize binary and data files as bitmaps, complementing a text viewer [76] C22. Timeline plots and icons arranged in a spiral, complementing log files [77] C23. Application of Chunking: collecting lower-level details in a single higher-level abstraction that can help understanding of a large problem [75] C24. Note taking [92]

Table 2.1: Elements of (A) cognitive models of binary RE, (B) concepts of cognitive theory, and (C) VR affordances/tools/techniques

2.6.2 Primary Themes

Through an affinity mapping process reviewing each end-to-end thread, we aligned related threads into overarching *themes* that are critical to the process of binary RE. In this alignment process, we concentrated on that we expect will have a significant impact on the task of *binary* RE; due to the lack of related research specifically in embodied immersion for binary RE, many of our sources focus on source code RE or software engineering in general, and while we believe many of those concepts may apply in the binary domain, here we've exercised subjective judgement to narrow the focus to those elements that most strongly apply. Clustering these threads revealed the themes we present here:

- Enhancing abductive iteration (hypothesis loop)
- Augmenting working memory
- Supporting information organization and discovery of important features

Figure 2.6 depicts the three themes and the most closely related elements from groups A, B, and C. These three themes touch upon every element from group A, cognitive modelling of binary RE, which is important because we want to find the most effective ways to augment as many of the cognitive model elements as possible. Moving into the group B elements of cognitive theory, we are more selective in our areas of concentration. Finally, we derive a central set of VR affordances from group C on which we will focus our efforts in future work. We will examine each theme more closely in the remainder of this section.

Enhancing abductive iteration (hypothesis loop)

One very prevalent theme in cognitive models of RE is the iterative pattern of sensemaking using abductive reasoning. In this analysis, that pattern is broken into four elements of abductive iteration: *setting goals/following plans*, *forming hypotheses*, *experimenting to test hypotheses*, and *updating what is known* [2], [34], [27], [28], [3], [4], [33], [30]. The cognitive model elements of *multi-level internal semantic representation* [1] and *heavy dependence on working memory* [1], [29] also broadly apply to this theme.

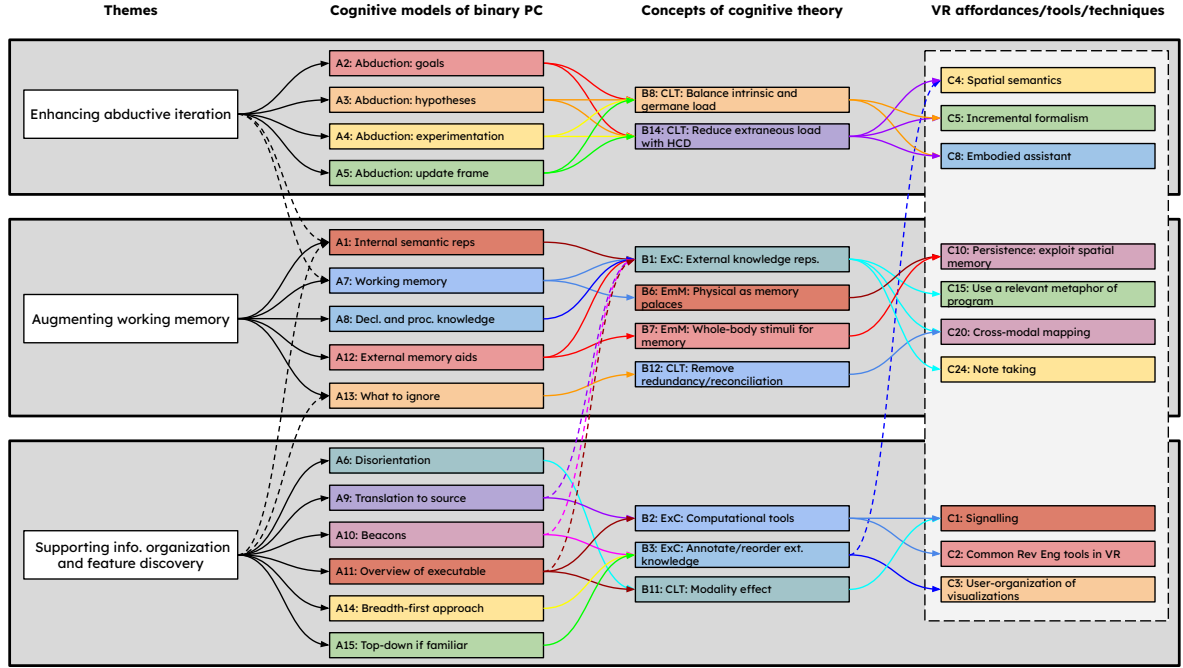


Figure 2.6: Primary themes for analysis with most closely-related elements; highlighted area indicates highest-priority VR affordances

The elements of cognitive theory most closely related to this theme include managing cognitive load in the iterative sensemaking loop through *balancing the intrinsic and germane loads* [52] and by *reducing extraneous load as much as possible through human-centered design* [55]. As this theme encompasses the overarching execution of the RE task, it has overlap with the other themes. For example, relevant elements from cognitive theory include those related to information organization and memory. In this theme, we narrow the focus specifically to facilitating the iterative loop and reserve examination of specific elements related to information and memory for later in this section.

In the iterative process, the prime opportunity for augmentation is in amplifying the expert's ability to identify and track progress toward goals and hypotheses. *What is the current goal? What is the current hypothesis? What evidence has been built? What are prior decision points to which we can return upon hitting a dead end?*

The immersive affordances most applicable include *incremental formalism* [58], in which the visualization of the iterative process evolves its structure as progress is made; *spatial semantics* [58], in which the spatial organization of the information provides semantic information; and the *embodied assistant* [71] that would (ideally) track the expert’s decision path and provide the interactive interface by which the expert can ask questions, make notes, etc.

Augmenting working memory

The limitations of working memory impact the effectiveness and efficiency with which analytical problems are solved, binary RE or otherwise. Closely-related elements from cognitive models of RE include *multi-level internal semantic representation* [1], *taxing working memory* [1], [29], *generation, storage, and retrieval of declarative and procedural knowledge* [2], *using external memory aids* [37], [36], and *determining what to ignore* [12].

The most closely-related elements of cognitive theory include *using external knowledge representations to reduce memory load* [42], [24] by offloading some knowledge that otherwise would be held in working memory, *employing memory palaces aided by physical objects or locations and leveraging whole-body stimuli to expedite storage and retrieval of memory* [50], and *removing redundancy of information presented in different modalities* [54] to streamline the intake of new information into working memory.

Following those threads from the cognitive model to cognitive theory leads us to a few considerations for immersive VR: *employing persistence to exploit spatial memory* [58] to remember information, *using physical metaphors* [87], [88], [90], [86], [89], [91] for the task’s information and operations, *cross-modal mapping* [61] to present information using multiple senses, and *note taking* [92] to offload working memory and share with collaborators. Additionally, reducing extraneous cognitive load, as described in the previous theme, can ensure working memory capacity is freed for solving the intrinsic problem.

Supporting information organization and feature discovery

Of the three themes, *supporting information organization and feature discovery* is the most context-dependent; while the themes of iterative abductive process and enhancements

to working memory can apply to almost any analytic problem, this theme is most tightly-integrated with the problem domain and its existing methods and tools.

The elements from cognitive models of binary RE most closely tied to this problem domain include *disorientation following execution paths* [29], *translating the binary back to source code* [33], *using or marking beacons* [27], [3], [4], *referring to an overview of the binary program* [4], and *approaching the task in a breadth-first and top-down manner, depending on the expert's familiarity* [35], [38]. Additional related elements, crossovers with the other two themes, include *multi-level internal semantic representations* [1] of the composition of the binary program and *determining what to ignore* [12] from the masses of data in the program and system.

Several cognitive theory elements support this theme. *Annotating and reordering/restructuring external representations of knowledge* [42], [24] addresses how experts take notes and organize process artifacts. *Using computational tools to make tasks easier* [42], [24] covers the use of current well-known reverse engineering tools to discover key characteristics of the binary in an automated or semi-automated fashion. The *modality effect* [54] can also be leveraged given multiple sources of information and interpretations of the same binary program, for example, to provide signalling cues. Finally, as a crossover element with the working memory theme, *using external knowledge representations to reduce memory load* [42], [24] acknowledges the need to use the external knowledge representations mentioned at the start of this paragraph.

Several VR affordances apply to this theme. First, *incorporate common reverse engineering tools* in a way that allows the expert to naturally interact with them from within the immersive environment; this task may require a novel interaction paradigm. Second, *signalling* [72], [73] will help the expert focus on what is most important; what is signalled may come from self-sourced notations, or it may come from cues inferred from the reverse engineering tools. *Spatial semantics* [58] applies here just as in the iterative abduction theme: the spatial organization of the data provides meaning itself, over and above the meaning of the data. Finally, *user organization of visualizations* [66] applies, especially in combination with the preceding element, where that organization in space is itself meaningful.

We note that there are three similar affordances in VR across the themes that may seem redundant, so we want to differentiate them. *Spatial semantics* exploits spatial position and representation to help the practitioner *understand*. *Persistence* exploits the same to help the practitioner *remember*. *User organization* exploits the same to help the user *express knowledge*. All three affordances are part of exploiting the immersive space to improve performance [57].

2.6.3 Section Summary

Three primary themes capture the salient aspects of cognitive models of binary RE. Following the threads that connect cognitive models to cognitive theory and then to VR affordances, we derived a set of the most relevant qualities for an immersive experience for binary RE:

- *Spatial semantics* [58], in which the spatial organization of the information provides semantic information
- *Incremental formalism* [58], in which the visualization of the iterative process evolves its structure as progress is made
- *Embodied assistant* [71] that (ideally) tracks the expert's decision path and provides the interactive interface by which the expert can ask questions, make notes, etc.
- *Employing persistence to exploit spatial memory* [58] to remember information
- *Meaningful and relatable metaphors* [87], [88], [90], [86], [89], [91] for the task's information and operations
- *Cross-modal mapping* [61] to present information using multiple senses
- *Note taking* [92] to offload working memory and share with collaborators
- *Signalling* [72], [73] focuses the expert's attention on potential key discoveries
- *Incorporate common reverse engineering tools* to leverage powerful tools already familiar to experts

- *User organization of visualizations* [66], in conjunction with spatial semantics, allows the expert to capture knowledge in the positioning of the data itself

In addition to these specific elements, we must consider overarching considerations for an immersive VR implementation. Any successful implementation will follow an *iterative human-centered design and development process*. It will incorporate automation as much as possible, including taking advantage of large language models. Additionally, it will be *adaptive* and self-tune the parameters of the immersive experience to give the expert user the most effective experience.

2.7 Conclusion

The process of binary RE is inherently complex and cognitively demanding, requiring very specialized expertise to perform it effectively. The problem is only getting harder as computer architectures become more varied and complex and binary obfuscation techniques become more sophisticated. Augmenting the cognitive process of binary RE is crucial to maintaining or improving the current level of effectiveness of experts performing this task.

To understand how we might augment the process, we surveyed prior work in three progressive groups: characteristics of mental or cognitive models of sensemaking in binary RE (answering RQ1), cognitive theory and applications to binary RE (answering RQ2), and cognitive augmentation of sensemaking using visualization and immersive technologies (answering RQ3). In our synthesis, we identified several common and salient elements in each group, assembled threads of related elements, and further mapped those threads to three primary themes (answering RQ4). First, reverse engineering can be framed as a process of abductive iteration, where hypotheses are formed, tested, and refined over time, which places a significant burden on working memory that is often mitigated through external aids and spatial structuring of information. Second, cognitive theories, such as cognitive load theory, external cognition, and embodied cognition, offer concrete principles for designing tools that support sensemaking, learning, and memory. Finally, prior work in immersive analytics demonstrates a range of VR affordances and techniques, such as spatial organization, physical navigation, contextual

persistence, and embodied interaction, which can be leveraged to reduce cognitive load and enhance exploratory reasoning in reverse engineering tasks.

2.7.1 Threats to Validity

While this review offers a cross-disciplinary synthesis of research relevant to immersive binary reverse engineering, several limitations must be acknowledged. First, the hybrid search strategy may have excluded relevant studies due to keyword mismatch, database limitations, or implicit biases in inclusion criteria. The relatively small body of directly related work also increases the risk of overgeneralizing from tangential fields such as immersive analytics or source-code visualization. Additionally, the thematic synthesis involves interpretive judgment across diverse domains. Finally, as this field evolves rapidly, some findings may be subject to change, highlighting the need for continued empirical validation and refinement of the proposed design implications.

Chapter 3

Design Process and System Architecture

To ensure that we build a system with features founded on qualities of human cognition and that is highly usable, we adopted a hybrid approach to our work. This approach combines Design Thinking (DT) [94] with Cognitive Systems Engineering (CSE) [11]. This chapter explains our iterative process and the resulting system architecture and features, which answers RQ5, *What design and architectural features are required in an immersive VR system to support static binary RE?* We used this process to define and execute three iterations to determine these features, starting with a Minimally Viable Product (MVP) and using informal feedback sessions to refine the system until it was ready for a formal user study described in Chapter 4.

3.1 Design Thinking and Cognitive Systems Engineering

We are applying the DT approach to our research phase-by-phase (discovery, definition, development, delivery) and merging CSE into that process. DT aims to improve the user experience based on design and testing and is often employed in problem solving, particularly in designing products. It leans on methods such as interviews, contextual inquiries, probes, and brainstorming to collect requirements and driven by empirical measurement during execution. As we are building a software system to provide a highly interactive and immersive experience to practitioners, DT is applicable with its history of success in building user-facing software. We adopted the British Design Council's Double Diamond [95] (see Figure 3.1) as our DT

approach. In the left diamond, we start with divergent thinking to explore the issue then use the discovered information to converge on a well-defined instance of the issue. In the right diamond, we first develop a wide range of potential designs, then narrow down to an implemented solution. We then iterate the right-hand diamond as we continue to pursue this line of research.

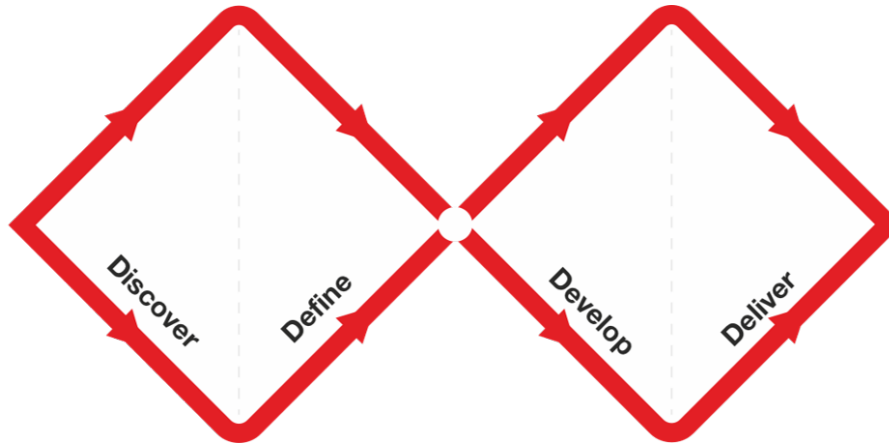


Figure 3.1: Double diamond.

CSE focuses on improving the joint human-machine system, particularly for complex systems like those used to support binary RE. Originating from cognitive psychology, it seeks to optimize systems through understanding and improving cognitive processes. As we described in Section 2.2, CSE has three steps: (1) identify the problem areas of the task; (2) understand the circumstances of the identified problems, focusing on potential causes and influences of those problems; and (3) pursue practical solutions to the identified problem areas [11]. Ultimately, CSE guides our investigation of the cognitive aspects of our work, while DT ensures our software system is highly usable in a practical sense. Figure 3.2 gives an overview of how we combine the two approaches into a hybrid process that we describe in more detail in the remainder of this section.

3.1.1 Discovery Phase

A primary component of CSE is Cognitive Task Analysis (CTA), which identifies the cognitive skills and mental demands of performing a task [96] and makes up a very significant part of our work in the discovery phase. CTA provides the foundational understanding upon which CSE would build an improved binary RE process. Our literature survey, particularly

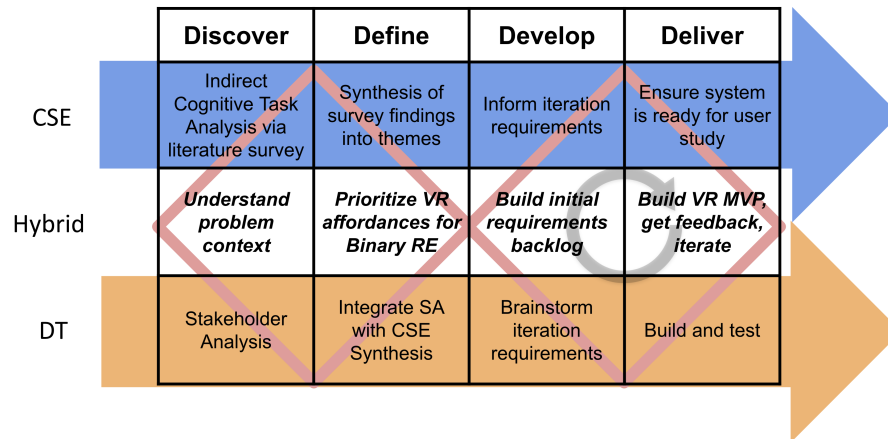


Figure 3.2: Hybrid process detail.

Section 2.4, forms an indirect and composite task analysis through its synthesis of prior task analyses of binary RE.

To augment the CTA, we engaged in a stakeholder analysis, which included three main elements as follows:

- Define the “As-Is” binary RE workflow:** We defined a notional practical workflow for binary static analysis supporting RE. It is broken into basic and advanced analysis and requires manually correlating the outputs of various tools. Basic static analysis includes steps such as reviewing the file header; identifying the type of file (architecture, format, etc.); reviewing strings (may need to be decrypted with another tool); reviewing symbols, function names, and library calls; and searching for embedded files or resources. Advanced static analysis includes disassembly and decompilation; code overview; reviewing the function call graph and function-level control flow graphs; pinpointing items of interest; adding comments or annotations; renaming functions (replacing auto-generated names with ones that have semantic meaning), disassembling at a micro scale to correct disassembly errors; and defining data structures. This workflow varies from practitioner to practitioner.
- Capture the impacts of “As-Is” in an empathy map:** Through observation and direct questions, we built an empathy map to understand how the current RE workflow affects practitioners. See Figure 3.3.

- **Perform a contextual inquiry to hypothesize “To-Be:”** Brainstorming with RE practitioners, we captured their desires for a more intuitive experience with RE tools. They identified needs in several areas: identifying and tracking toward goals and hypotheses (designating them and tracking progress); enhancing working memory through external aids (e.g., a notepad); eliminating extraneous cognitive load so that working memory capacity can be used for the intrinsic problem instead of overhead workload (e.g., masking out what can be ignored; employing automation); identifying, marking, and tracking beacons; and providing an expedited way to get an overview of the complete binary program while still providing tools to drill down.



Figure 3.3: Empathy map.

3.1.2 Definition Phase

The Discovery phase revealed a wide array of considerations. The purpose of the Definition phase in the Double Diamond design process is to systematically distill those findings into a realistic initial set of requirements. Again, we merge DT and CSE approaches. The CSE approach was described in Chapter 2, in particular, the identification of threads and themes in Section 2.6 that highlighted core cognitive concerns and affordances to address them. From the DT perspective, findings from the workflow analysis, empathy mapping, and contextual inquiry

provided user-centered insights into tool friction, navigation breakdowns, and opportunities for intuitive spatial layout.

We then prioritized the affordances that would provide the largest returns with the smallest development effort to facilitate early feedback. This process did not define how they would be implemented; that will be addressed in the development phase described next. We identified these affordances as initial system requirements: spatial semantics, incremental formalism, persistence, metaphors, signalling, user organization of visualizations, notepad, and of course, incorporating common reverse engineering tools. As stretch goals, we also considered an embodied assistant and cross-modal mapping.

Following this process ensured that our early development effort focused not just on VR novelties, but on aligning system capabilities with known cognitive needs and usability principles.

3.1.3 Development → Delivery Phase Iterations

Once we defined the initial scope of work, we moved to the right-hand diamond of the double diamond approach encompassing development and delivery. We are executing multiple iterations of the “Develop → Deliver” diamond. These iterations implement new or updated features and are driven by user feedback. Their purpose is to deliver a system that is reliable to ensure valid and repeatable results, user-friendly to reduce extraneous friction points, and relevant to the RE task. The first iteration encapsulates a large startup effort to build a “minimally viable product” (MVP) as a starting point for further minor iterations. These iterations led up to a formal user study to gauge effectiveness of our implemented affordances.

3.2 System Architecture

Our iterative Development → Delivery phase activities centered around building a platform, Cognitive Binary Reverse Engineering (CogBRE), collecting feedback, and improving

its features and user experience. CogBRE integrates the capabilities of multiple common reverse engineering tools into an immersive environment using virtual reality. This platform introduces novel ways to work in this domain with the intention of reducing cognitive load.

In order to test our theories on the application of VR affordances to the binary RE process, we built CogBRE to allow implementing a wide range of relevant capabilities in an immersive environment. We have three primary design objectives for CogBRE:

- **Leverage existing reverse engineering capabilities:** There are many high-quality tools that offer means to control their functionality through automated means, such as command line scripting or application programming interfaces (APIs). We want to reuse these tools rather than implement redundant capabilities.
- **Support multiple VR, AR, MR clients:** Although our first focus is on a fully immersive VR environment, we want to keep the system sufficiently modular in order to support other types of clients in the future, and to support multiple users.
- **Be comfortable to use for extended periods:** The process of binary RE can take hours. The system needs to minimize strain, fatigue, and simulator sickness as much as possible.

In working toward these objectives, we designed a set of modular capabilities. Figure 3.4 shows the high-level architecture of the system. The remainder of this section will provide additional details of the major system components.

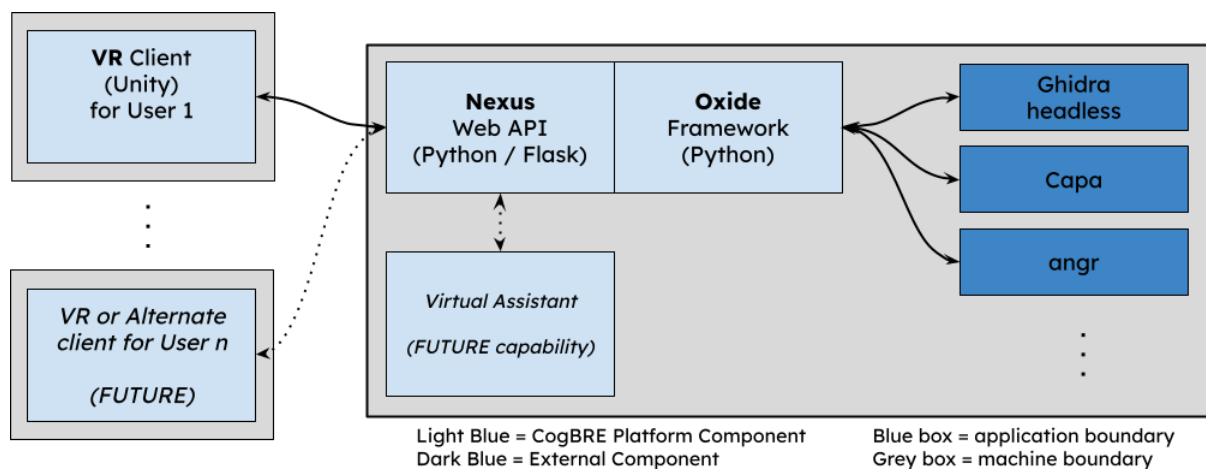


Figure 3.4: CogBRE Platform Architecture

3.2.1 Oxide

Oxide is a flexible tool for performing analysis of binary programs [97]. The design is modular, allowing the integration of multiple third-party tools along with custom analysis modules. For example, Oxide might pull disassembly from Ghidra [7], semantic features from Capa [98], and a control-flow graph from angr [99]. Since each of these tools handle loading differently, the Oxide modules map the address space into a common space based on file offsets. This allows us to easily compare and integrate results. Building on top of Oxide provides the following benefits:

- **Modularity and Extensibility:** Oxide allows us to integrate research happening in parallel, within our lab, other universities, as well as in industry.
- **Domain Separation:** One of Oxide’s goals is to allow researchers who specialize in a specific domain to focus on implemented features relevant to their domain.
- **Low Framework Overhead:** Oxide returns results in a very simple JavaScript Object Notation (JSON) or Python dictionary-based format that assumes very little and allows integration of a wide range of results.
- **Caching:** In order to keep a VR system responsive, we needed to minimize the time spent waiting for binary analysis results. Oxide aggressively caches any analysis that has been previously performed.

For this project, we implemented a custom interface for Oxide called the Nexus that facilitates interactivity between the VR client and the Oxide back-end. Oxide has a public code repository at <https://github.com/Program-Understanding/oxide> .

3.2.2 Nexus

The Nexus is the broker between the client systems and data sources. Implemented as a lightweight Python module, it provides a basic RESTful web API to clients using Flask 1.1. Its runtime instance hosts Oxide, which in turn connects with various data sources. Upon a

client API call, it invokes appropriate method(s) in Oxide, processes the return values into a JSON payload if necessary, and returns that to the client. The web API design pattern allows a wide variety of clients to easily connect, and Nexus provides a rudimentary mechanism to track multiple users. This pattern also makes it easy to place the intense workload of reverse engineering tools on a different host than the also-intense workload of rendering the VR display. Nexus has a public code repository at <https://github.com/AugCogVR/cogbre-nexus>.

3.2.3 VR Client

Of many potential client types, we first implemented a VR client, which is required to test our hypotheses about the affordances of an immersive environment. We provide more details on the design in several categories, as follows:

- **Technology:** The VR client is developed in Unity 2021.3.4f1 and runs on common desktop operating systems. We use Microsoft's Mixed Reality Toolkit 2 to provide many of the VR components. Our development targets SteamVR devices; in particular, we are using the HTC Vive Pro 2 head-mounted display due to its display resolution and field of view.
- **Software Architecture:** The architecture adopts the common pattern of a singleton GameManager that maintains references to instances of important objects in the environment and controls environmental conditions on update. New instances of virtual objects are created and destroyed based on user actions. Figure 3.5 illustrates the class hierarchy.
- **Information Architecture:** The VR application mirrors Oxide's hierarchy of collections (of files), files (primarily binary programs), functions, and basic blocks through its own hierarchy of those respective classes. Additionally, it provides access to all collections that are created in Oxide and all binary files that are imported into Oxide, via Nexus. As more information about a binary or a function is pulled via Nexus (for example, header information, disassembly, or decompilation), the corresponding VR object instances are progressively augmented.

- **Performance Considerations:** In an effort to remain performant and engaging for users, VR pulls data from Nexus “on demand” as the user requests it. We also fully leverage Oxide’s intelligent caching capability described in Section 3.2.1, maintaining a high level of interactivity in the application.
- **Quality-of-Life Considerations:** Because a binary RE practitioner may spend an extended period of time in the VR application, we’ve taken steps in our design to increase presence and reduce cybersickness [100]. The environment is at room scale (implied) and static except when objects are directly manipulated by the user, and for now, locomotion is only via physical user movement. Additionally, we maintain a responsive frame rate even during intensive operations or long-running transactions using Unity’s Coroutines for enumerable tasks and C# Asynchronous Tasks for monolithic transactions to minimize the computation during each frame.

The VR Client has a public code repository at <https://github.com/AugCogVR/cogbre-vr>.

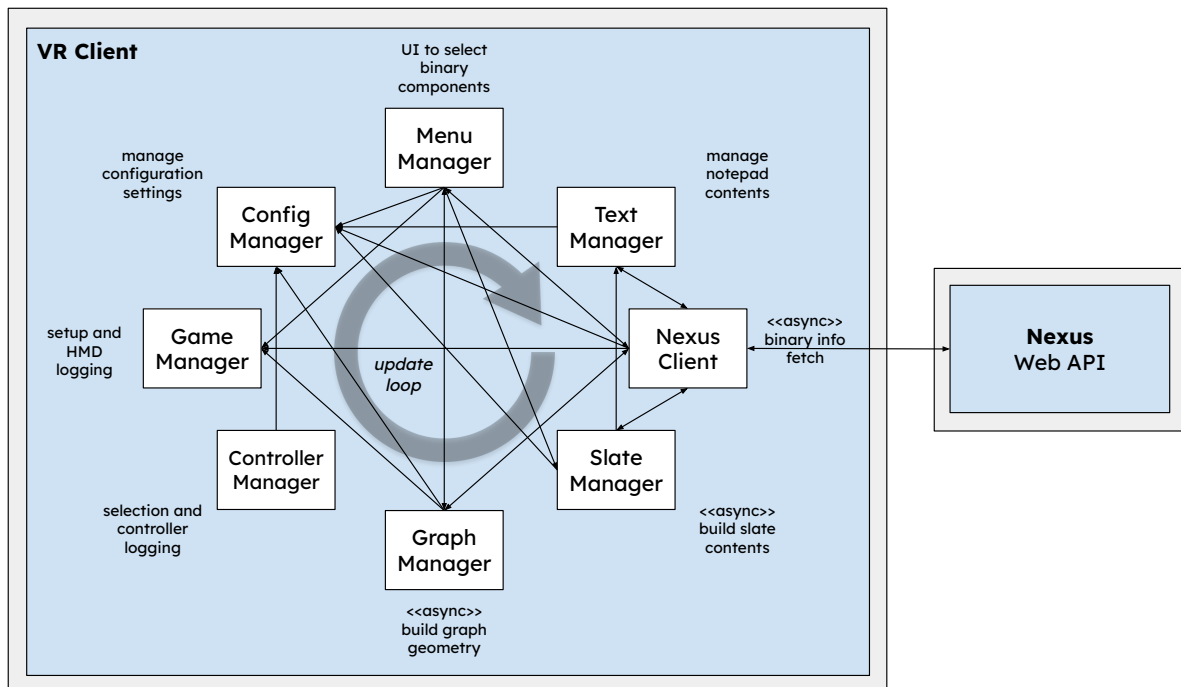


Figure 3.5: The VR Client, as a Unity application, consists of modular “managers” extending the MonoBehaviour base class that run in the main thread; some of them spawn threads to accomplish long-running tasks with minimal impact to the main thread, improving interactivity and usability.

With these system components (Oxide, Nexus, and VR) in place, we embarked upon an iterative user-centered design process.

3.3 Iteration 1: Minimally Viable Product

Our discovery and definition process revealed three primary areas of opportunity to improve the cognitive circumstances of binary RE practitioners: *Enhancing abductive iteration (hypothesis loop)*, *Augmenting working memory*, and *Supporting information organization and feature discovery*. That work also identified several affordances of VR that can influence cognition related to these themes.

3.3.1 Feature Implementation

Our discovery and definition process also included a stakeholder analysis. Contextual inquiry of binary RE practitioners revealed common elements of their practical binary RE workflows using contemporary reverse engineering tools as listed in Section 3.2.1 as well as ideas about how the process can be improved. Finally, an empathy mapping exercise captured additional practitioner insights. Based on the outputs of the definition phase, we determined the features to implement in our MVP. The table in Figure 3.6 traces the three themes to their mostly closely-related VR affordances to the features implemented using those affordances.

The features we chose to implement first heavily focus on the use of the vast 3D space, strongly influenced by Virginia Tech’s *Immersive Space to Think* research [57]. The MVP provides practitioners somewhat traditional visualizations of binary files in the 3D space by employing two primary visualizations:

- **Graphs** visualize the elements and flows at two levels. *Call graphs* display functions of a binary and the flows between them. *Control Flow Graphs* display basic blocks of a function and flows between them. The graphs can be rendered via the hierarchical Sugiyama method [101] or with a force-directed method. Users can select functions for further analysis.

Theme	Affordance	MVP Implementation
1. Enhancing abductive iteration (hypothesis loop)	C4. Spatial semantics (understanding)	Hierarchical (Sugiyama) graphs of function calls and control flow
	C5. Incremental formalism	User can manually explore Call Graph → CFG → Code
	C8. Embodied assistant	(not yet)
2. Augmenting working memory	C10. Persistence: exploit spatial memory (remembering)	Environment provides basic tools for user-managed placement
	C15. Use a relevant metaphor of program	Basic hierarchical (Sugiyama) graphs only
	C20. Cross-modal mapping	(not yet)
	C24. Note taking	(not yet)
3. Supporting information organization and feature discovery	C1. Signalling	(not yet)
	C2. Common Rev Eng tools	Exploit via chain of Nexus → Oxide → Ghidra
	C3. User-organization of visualizations (expressing)	Create, Move, Delete slates and graphs arbitrarily in 3D space

Figure 3.6: Iteration 1 (MVP) Features

- **Slates** display scrollable outputs of tools brokered by Oxide and Nexus, such as header information, strings, function disassemblies, and function decompilations.

Users can create, reposition, resize, and delete slates and graphs without limitation in the 3D space. Additionally, a basic slate-based menu of scrolling lists allows users to select the Oxide collection, binary, and function to analyze. Users can switch between collections, binaries, and functions without impact to visualizations previously generated and placed in the space. Figures 3.7 through 3.10 provide additional screenshots of a user session in an early version of the system.

3.3.2 Feedback

We held sessions to solicit feedback from RE practitioners on the current system implementation. These sessions do not constitute a formal effectiveness study, which we plan to conduct later this year, but they give us valuable early feedback. This feedback was obtained from five different individuals associated with our research group (four graduate students and one undergraduate), having moderate or higher levels of experience with binary RE, and no to moderate experience with VR.

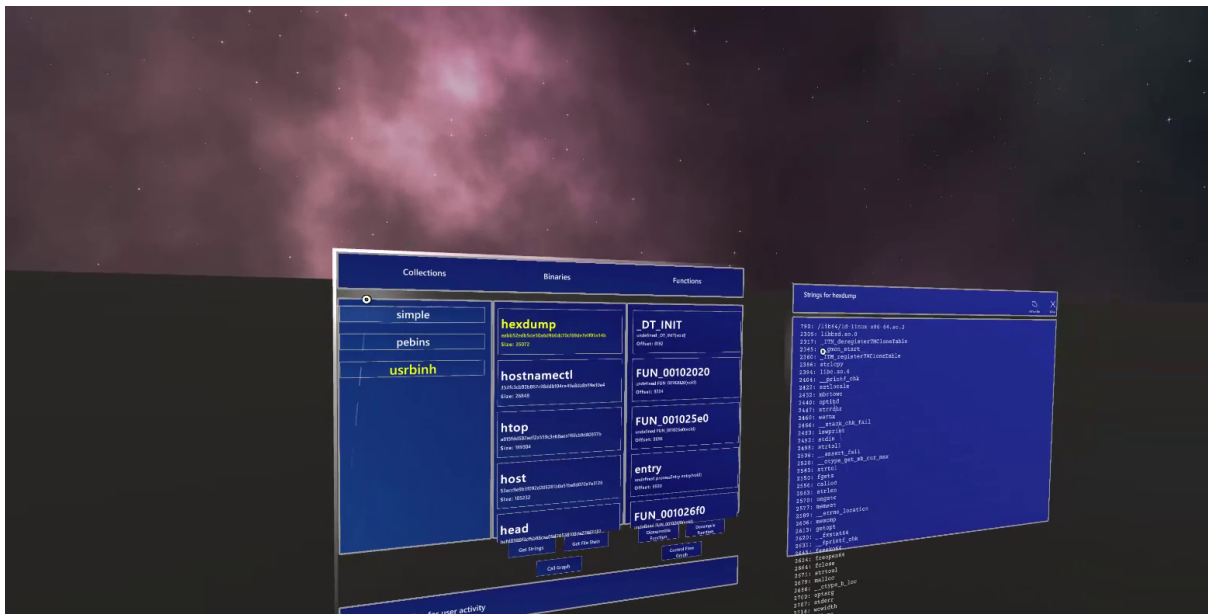


Figure 3.7: Iteration 1 screenshot: the main menu and a strings slate; text scrolling in slates was not yet implemented.

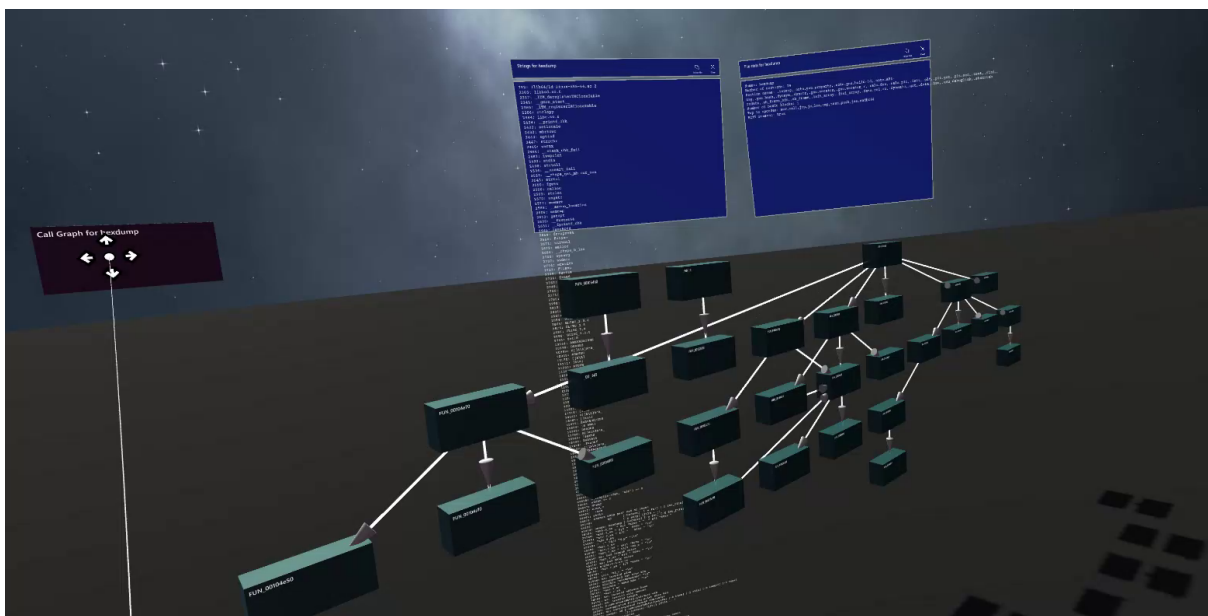


Figure 3.8: Iteration 1 screenshot: a call graph and informational slates.

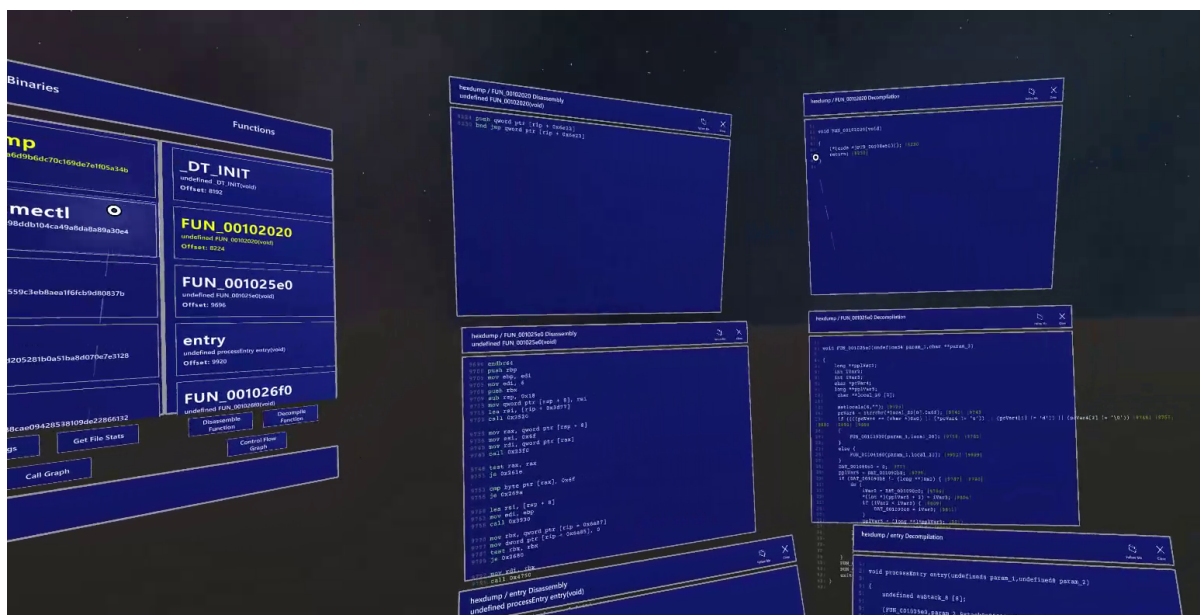


Figure 3.9: Iteration 1 screenshot: function disassemblies and decompilations.



Figure 3.10: Iteration 1 screenshot: decompilation slates and control flow graph.

We began our feedback sessions with a brief demonstration of the current capabilities of our program; users were then free to use the program as they saw fit, while we observed their experience through the lens of empathetic questions: **What did the users do/say/think/feel?** This open-ended observation yielded these highlights: Users found the system easier to use in person than they expected based on prior videos and still frames of the system in action, and they needed little-to-no direction after donning the HMD and controllers. They remarked on the flexibility of the system compared to other reverse engineering tools discussed earlier, such as being able to analyze multiple binary files in the same environment. One individual who often feels nausea in VR did not in this case. Finally, users often mentioned illegibility of the smaller fonts; this is likely because the system was tuned for one particular HMD and, due to unfortunate circumstances, a different model with lower resolution was substituted for the feedback sessions.

Additionally, we employed a direct line of questioning to obtain explicit feedback. Summations of the feedback received are as follows:

- **Would you use this tool as-is when trying to comprehend a binary program?** This question yielded mixed results; most (but not all) users said they would not use this program as is. Negative responders primarily cited poor text legibility (addressed above) and lack of functionality as to why they would not use this version of the program.
- **What is better in the VR environment than tools and processes you usually use?** Users stated that obtaining information about the binary was significantly more intuitive than state-of-the-art tools discussed earlier. Furthermore, users overwhelmingly cited that the vast spatial real-estate for management and comparison of different tasks supersedes the comparison capabilities present in the aforementioned tools.
- **What is worse?** In contrast, users lamented that the system offered no means to make annotations and notes in the environment. Additionally, users stated that they wanted the program to match the functionality present in state of the art tools (e.g., function renaming, structure definition, library call identification, etc.).

- **What are minor improvements you want to see in this tool?** Users predominantly stated that they wanted a means to integrate each abstraction into a larger unified graph, primarily through a “find all instances” function that takes in a selection of disassembly, and finds all instances of the assembly throughout each representation. Furthermore, some users wanted a means of condensing the information down into collapsible graphs that would progressively grow as they dived deeper into the binary.
- **With those minor improvements, would you use this tool when trying to comprehend a binary program?** Users unanimously stated that they would use this program with the aforementioned minor improvements, with the exception of one inconclusive answer. Most users stated that they would apply this technology in the context of solving very basic reverse engineering tasks if these improvements were implemented.
- **What are major improvements you want to see?** Users were widely varied in this inquiry. Most wanted to see advanced integration with tools such as angr, and more robust comparisons between disassembly/decompilation across different functions.
- **With those major improvements, would you use this tool when trying to comprehend a binary program?** Users predominantly stated that they would use CogBRE under the stipulation of improvement.

We used the feedback from Iteration 1 to improve the operation of our system in Iteration 2 as described in the next section.

3.4 Iteration 2

3.4.1 Feature Implementation

In Iteration 2, we implemented several features to elicit additional feedback from practitioners. Figure 3.11 shows the updated mapping from theme to affordance to implementation. However, some changes were simply quality-of-life improvements that transcended specific affordances: Participants felt the relatively featureless environment was a missed opportunity

to include spatial reference indicators, so we updated the environment to a home office (see Figure 3.12), and they found the blue slates visually jarring, so we based the new color scheme on neutral/gray colors (see Figure 3.13). We also added a radial menu to allow quick access to primary functions (see Figure 3.14) and a magnifying glass/telescope so users can view fine details without manipulating an object’s position or scale (see Figure 3.15).

We also implemented or updated several affordances. The notepad and on-screen keyboard were most anticipated, allowing users to directly input text, providing in-situ memory augmentation (see Figure 3.16). Users were also provided the capability to select text in slates and copy it to the notepad. We also implemented signals/beacons in the function call graphs by importing and displaying capability findings using the Capa tool [98] (see Figure 3.17).

Theme	Affordance	MVP Implementation
1. Enhancing abductive iteration (hypothesis loop)	C4. Spatial semantics (understanding)	Hierarchical (Sugiyama) graphs of function calls and control flow
	C5. Incremental formalism	User can manually explore Call Graph → CFG → Code
	C8. Embodied assistant	<i>(not yet)</i>
2. Augmenting working memory	C10. Persistence: exploit spatial memory (remembering)	Environment provides basic tools for user-managed placement
	C15. Use a relevant metaphor of program	Basic hierarchical (Sugiyama) graphs only
	C20. Cross-modal mapping	<i>(not yet)</i>
	C24. Note taking	Notepad
3. Supporting information organization and feature discovery	C1. Signalling	Call graph node colors and markings indicate Capa findings
	C2. Common Rev Eng tools	Exploit via chain of Nexus → Oxide → Ghidra
	C3. User-organization of visualizations (expressing)	Create, Move, Delete slates and graphs arbitrarily in 3D space

Figure 3.11: Iteration 2 Features: Updates are in bold (note taking and signalling).

3.4.2 Feedback

Following the first round of feedback and associated feature addition, we conducted a second round of user feedback collection. These sessions involved five individuals from our research group (only one repeat from the first round of feedback) with varying experience in



Figure 3.14: Iteration 2 screenshot: the radial menu allows users to enable and disable the main menu, notepad, and magnifying glass from the touchpad of the VR controller.



Figure 3.15: Iteration 2 screenshot: as an alternative to panning and zooming a large graph, the user can use the magnifying glass.

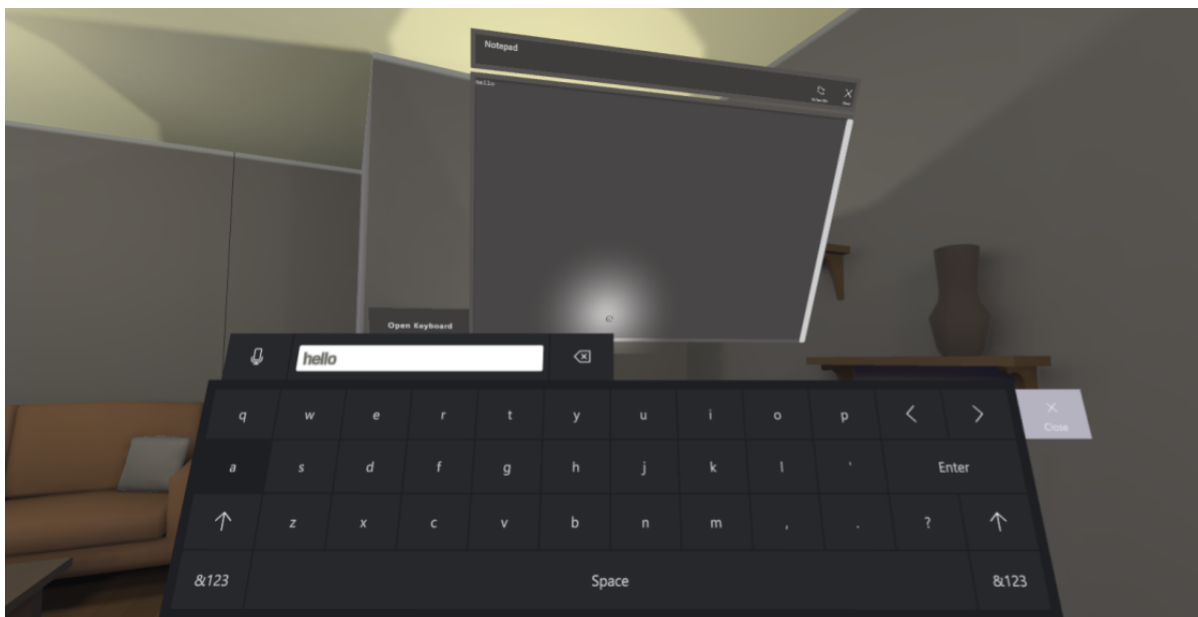


Figure 3.16: Iteration 2 screenshot: the new notepad and on-screen keyboard; in this iteration, the keyboard was severable from the notepad.

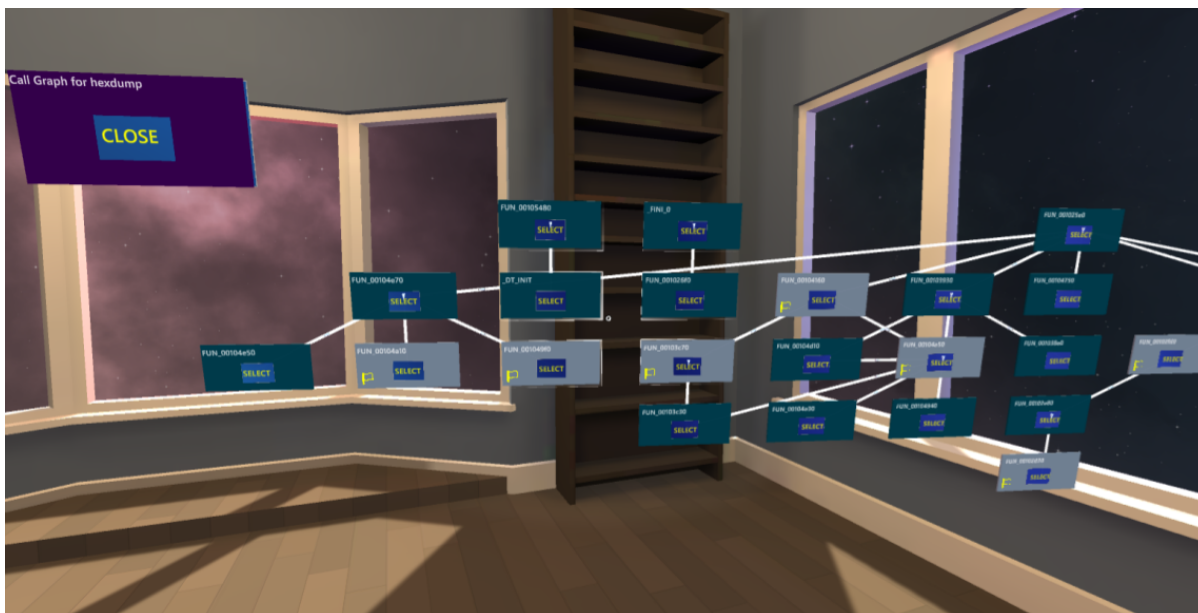


Figure 3.17: Iteration 2 screenshot: call graph functions with Capa findings are shown with visual distinction.

reverse engineering and VR, including at least one participant with no prior VR exposure. Users were encouraged to explore the system freely after a brief demonstration.

As before, we observed their experience through empathetic prompts: **What did the users do/say/think/feel?** Participants engaged deeply with the system: they opened and arranged call graphs, CFGs, disassembly and decompilation slates, and notepad windows; explored multiple binaries; and experimented with features such as copy-paste and magnification. One user filled the entire space with visual artifacts. Users physically moved throughout the space, interacted with spatial layout features, and tested interface affordances like text selection and resizing. They generally expressed satisfaction with the interaction model and praised the immersive spatial layout, but they also highlighted several areas of difficulty. Most users remarked on text legibility issues: white-on-gray slate text, small font sizes, and occasional stuttering when rendering long texts on slates. Others commented on environmental polish, such as “flashlight” flickering of the pointer selection indicators on slates (traced to Z-buffer conflicts and resolved) and UI elements (e.g., menus and graph close buttons) that were difficult to read or operate. No user reported motion sickness, although one user new to VR felt momentarily dizzy after a longer session.

We again asked our direct questions to the users and summarize their responses as follows:

- **Would you use this tool as-is when trying to comprehend a binary program?** Most participants stated that they would use the tool in its current state for simpler reverse engineering tasks. One participant explicitly said they would not, due to interaction and legibility concerns. Others qualified their “yes” responses by noting that use for more complex programs would be impractical without additional organizational tools and performance enhancements.
- **What is better in the VR environment than tools and processes you usually use?** Participants consistently emphasized the benefit of spatial organization. They found it intuitive and powerful to position and group artifacts throughout the virtual environment. Several noted that this obviated the need for screen or window switching, which they

found cumbersome in tools like Ghidra or IDA. One user described being able to “shove things aside” rather than minimizing or tabbing through windows.

- **What is worse?** In contrast, the most frequently cited drawbacks were tied to input precision, usability, and performance. Difficulties selecting text, resizing slates, and operating the virtual keyboard were common. Flickering visual elements, occasional interface bugs (such as a scrollbar stuck at 100%), and perceptual discomfort with very large graphs were also noted. As demonstrated in Figure 3.18, it was difficult for users to find a suitable scale for large graphs, balancing the ability to see the entire graph in the enclosed space with the ability to read the text in the nodes. Several users also mentioned that the system lacked interactivity common in RE tools, such as clicking on a function offset to navigate to another function or renaming identifiers.
- **What are minor improvements you want to see in this tool?** Participants offered a range of low-effort improvement suggestions, such as improving text readability by increasing font size and using a higher-contrast background; allowing dynamic resizing of slates based on text content; fix flickering effects described previously; implementing text scroll behavior with the thumb pad; improving text selection precision; providing visible feedback for actions such as selecting a function in the call graph; wrapping wide graphs around the user rather than requiring the user to rotate them manually; and optimizing performance when rendering dense or large visualizations.
- **With those minor improvements, would you use this tool when trying to comprehend a binary program?** Users unanimously responded that they would use the tool if the above minor improvements were made. They noted that even basic RE tasks would become more accessible, and several felt the tool could be integrated into their workflow for small- to medium-sized binaries.
- **What are major improvements you want to see?** Users offer the following suggestions for major enhancements: transitioning call graphs to a fully 3D representation rather than projecting 2D layouts in 3D space; introducing semantic interaction capabilities, such as

clicking offsets in the control flow graph to navigate; renaming symbols; enabling environmental segmentation, such as separate rooms or zones for different analysis contexts; creating visual linkages between related elements (e.g., linking offsets across graphs); enhancing input options for better ergonomics, such as gesture-based interaction or swipe input for the keyboard; integrating external media playback (music streaming); and offering a more open working environment such as a forest or warehouse.

- **With those major improvements, would you use this tool when trying to comprehend a binary program?** All users responded affirmatively. With the major improvements implemented, participants indicated that they would use the system for a broader range of tasks and found its conceptual direction promising. One participant, despite being new to VR, stated they would be excited to return to the tool if the environment were refined to support complex workflows.

We used the feedback from Iteration 2 to improve the operation of our system in Iteration 3 as described in the next section.



Figure 3.18: Iteration 2 screenshot: large graphs exceed space available in the new environment.

3.5 Iteration 3

Based on the feedback from Iteration 2, we made primarily quality-of-life improvements in several areas in preparation for the user study. We developed an environment modeled after a forest clearing, providing openness for large graphs and spatial arrangements of artifacts, but still somewhat closed (tree line), and containing various objects as spatial references (see Figure 3.19). We also improved text legibility, tweaking font size, colors, and lighting (see Figure 3.20). Two other minor improvements include clear visual indication of the grabbable area around each graph (translucent rectangle, see Figure 3.21) and attaching the keyboard permanently to the notepad and giving it direct control of notepad contents (see Figure 3.22). We did not collect informal feedback for this iteration as we were ready to proceed to the formal user study described in the next chapter.

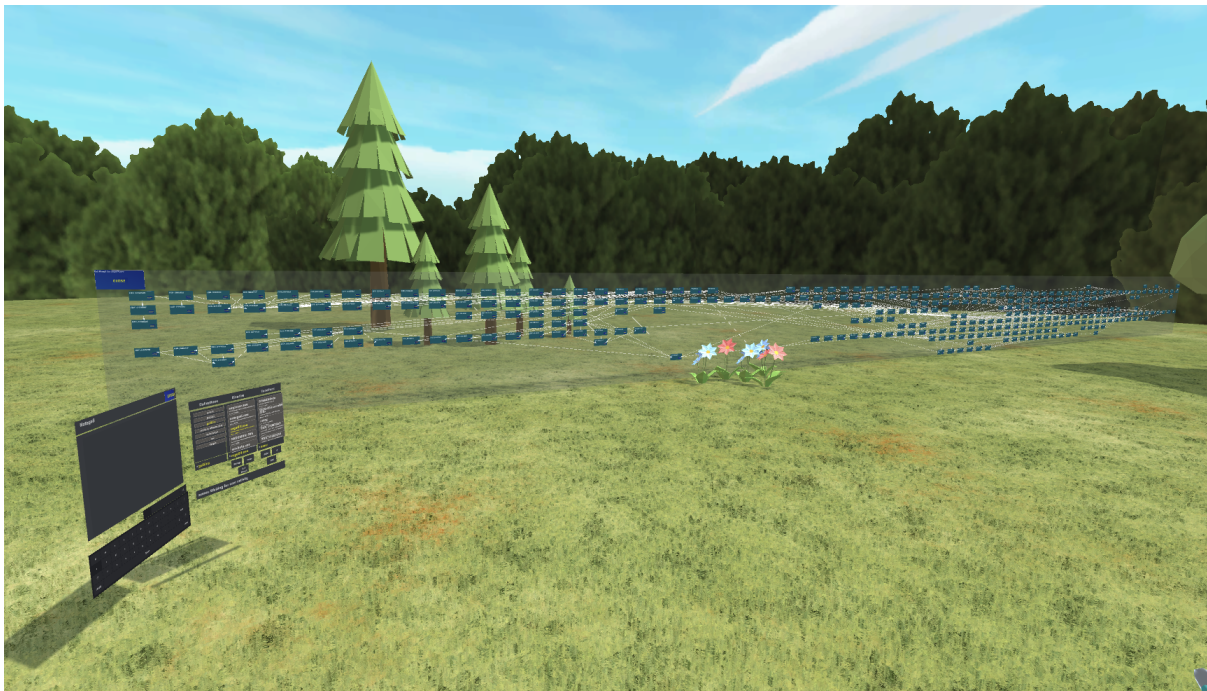


Figure 3.19: Iteration 3 screenshot: a large call graph (for Windows XP REGEDIT.EXE) fits easily in this environment, though the user still needs to pan/zoom/magnify as needed.



Figure 3.20: Iteration 3 screenshot: main menu and a text slate in the new open environment and text improvements.



Figure 3.21: Iteration 3 screenshot: a call graph for a simple binary with a control flow graph for one of its functions; darkened area indicates the grabbable area of the graph.

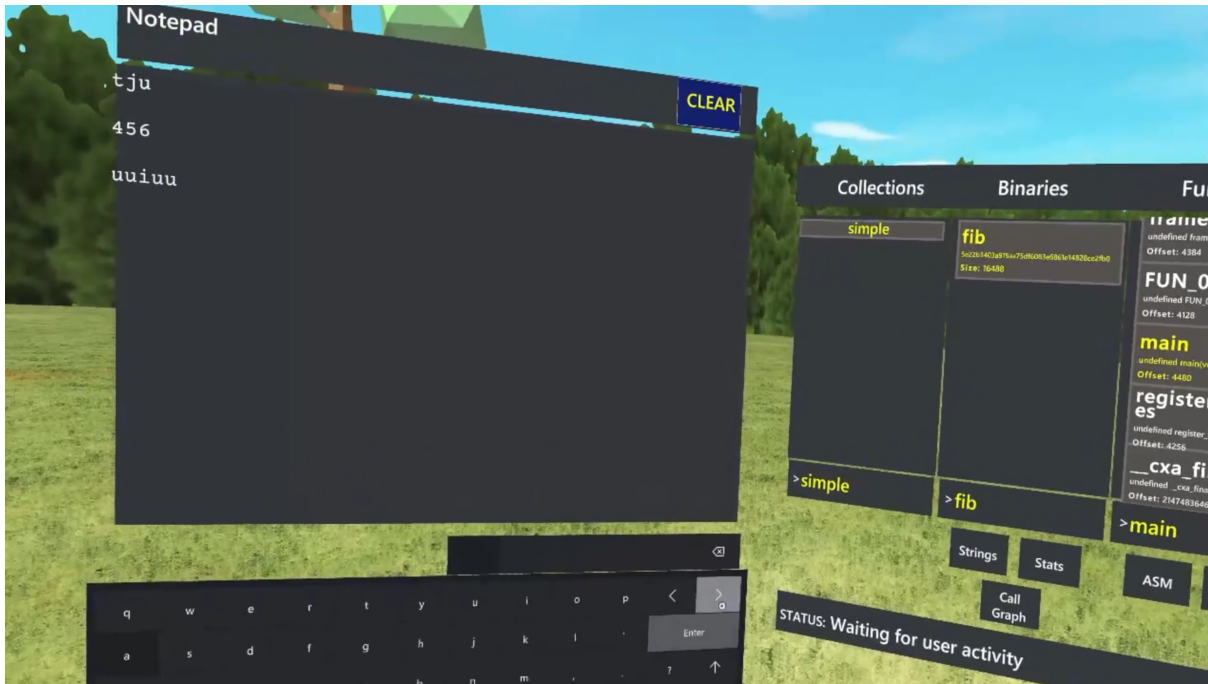


Figure 3.22: Iteration 3 screenshot: updated notepad with attached keyboard that provides direct control of the notepad text area as most users expect.

3.6 Conclusion

This chapter addresses RQ5: *What design and architectural features are required in an immersive VR system to support static binary RE?* Through a hybrid application of DT and CSE, we derived a set of affordances and system requirements grounded in user needs, cognitive theory, and the constraints of reverse engineering tasks. We allowed the design to emerge and evolve iteratively through discovery, definition, and feedback-driven iterations of development and delivery.

The initial set of requirements focused on spatial semantics, user-driven layout, visual signaling, persistent organization, and support for incremental formalism. Each of these features maps to specific cognitive or usability challenges and is supported by either observed user needs or theoretical principles ultimately related to sensemaking under uncertainty.

Architecturally, the system was designed to support modular integration of visualization, interaction, and RE toolchain outputs, enabling scalable development of a complex feature set refined by real-world feedback. These design decisions emerged deliberately in the execution of our hybrid process, resulting in a system ready for empirical evaluation in the next chapter.

Chapter 4

User Study

By our third development iteration, we had confidence to initiate a formal user study to measure how varying levels of spatial affordances in immersive VR impact effectiveness in performing binary reverse engineering tasks. This chapter describes our study; its execution and analysis address RQ6, *How do binary RE practitioners employ the affordances of embodied immersion in this system?* and RQ7, *Does this system improve outcomes in static binary RE tasks?.* The findings contribute to broader questions about embodied interaction and cognitive support in VR-based software analysis tools.

4.1 Study Design

We chose to frame the study as a Reverse Engineering Challenge in which we provide a purpose-built binary program for participants to analyze in one of three conditions. We asked a set of questions that require the participant to perform RE tasks on the binary program using tools we provide. These questions increase in difficulty and build on each other (scaffolded). Each participant completed the RE challenge under one condition while their performance, cognitive load, subjective experience, and interaction behaviors were recorded for later analysis.

4.1.1 Hypotheses

We guided our study to test four hypotheses, each grounded in theories of embodied cognition, spatial interaction, and cognitive offloading in immersive environments. The hypotheses reflect both performance outcomes and behavioral patterns that may emerge from participants' use of the system in the three experimental conditions. Our hypotheses include:

- **Hypothesis 1:** Participants will have better task completion metrics with certain levels of affordances in VR.

Rationale: Affordances that support spatial layout and flexible navigation are hypothesized to reduce working memory load and improve mental model formation, which should lead to more accurate and efficient performance on the reverse engineering task, particularly for questions requiring the integration of multi-step logic.

- **Hypothesis 2:** Cognitive load will vary by level of affordances in VR.

Rationale: Enhanced spatial affordances such as free placement, persistent object positioning, and navigable graph hierarchies may allow users to offload cognitive effort onto the environment. In contrast, fixed-layout conditions (basic VR and desktop) may require greater internal cognitive tracking of intermediate results.

- **Hypothesis 3:** Participants will rate certain affordances in VR (e.g., user organization, spatial memory, signalling) higher than others in terms of usefulness.

Rationale: As participants use the system, they may find that some features contribute more meaningfully to the task (e.g., affecting comprehension and recall) than others.

- **Hypothesis 4:** Participants who have better performance in the immersive environments will have used more of the VR space.

Rationale: Embodied cognition theory suggests that externalizing and organizing information in space can reduce cognitive load and enhance problem-solving effectiveness. Spatial telemetry data (in VR conditions) enables analysis of whether participants who use more of the available workspace achieve higher accuracy or efficiency.

Each hypothesis corresponds to specific elements of the system design, participant interaction data, and collected survey instruments.

4.1.2 Conditions

Participants were assigned to one of three experimental conditions (we used a between-subjects design; more details in Section 4.1.3): a desktop interface (web-based application), a basic VR environment, or an enhanced VR environment. All conditions provided access to a baseline set of tools, and the enhanced VR environment offered additional affordances.

Desktop (Web) Condition

The desktop condition used a custom web-based interface built with React and Flask. This interface connected to the same back-end system as the VR environments (see Figure 4.1) and exposed a subset of features suitable for static binary analysis. The user interface is shown in Figure 4.2. Users could navigate a main menu to select a binary and view program details (file statistics, strings, disassembly, and decompilation of functions) in text windows that appeared within a horizontally scrollable panel. A simple notepad feature was included to support annotation. Users could also view function call graphs in a separate scroll panel. The web interface was designed to mimic a minimal reverse engineering environment while maintaining feature parity with the VR Basic condition (described next). It provided no spatial manipulation capabilities beyond linear scrolling. The Web Client has a public code repository at <https://github.com/AugCogVR/cogbre-web>.

VR Basic Condition

The VR Basic environment provided an immersive, room-scale interface for reverse engineering, rendered within a forest clearing scene designed to support spatial orientation (via distinctive landmarks) while avoiding visual clutter. Objects in this environment were not subject to gravity or collisions, so they could be arbitrarily positioned anywhere by the participant.

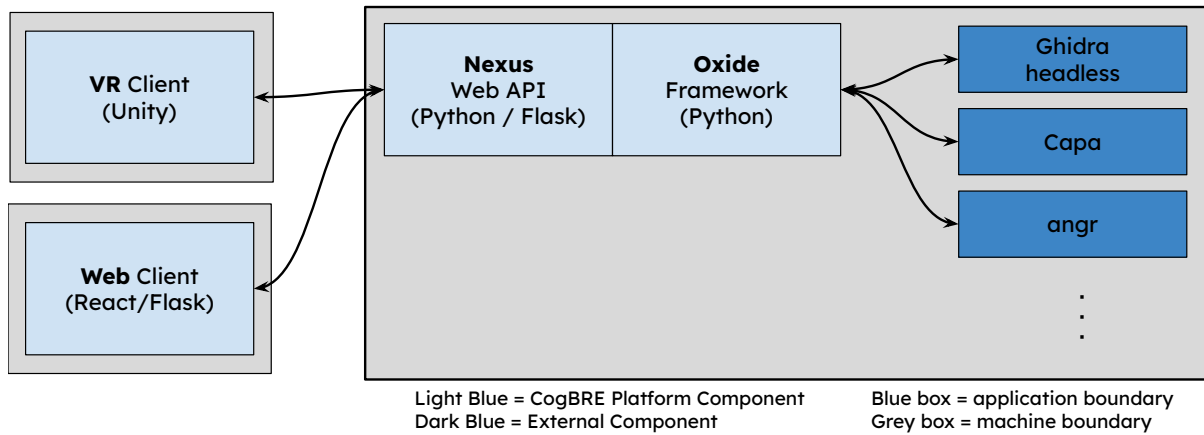


Figure 4.1: Web Client added to the CogBRE architecture.

This condition included a head-mounted stereoscopic display (HTC Vive Pro 2) and two hand-held controllers for interaction. The VR environment employed a natural movement locomotion model, where users physically moved within a tracked space and their motion was mapped 1:1 to virtual movement. This design choice was based both on requirements (it was sufficient for the task) and on prior research showing that physical movement increases presence and induces significantly less cybersickness than artificial locomotion methods such as flying, teleportation, or joystick-based navigation [102, 103]; the tight coupling between vestibular and visual cues in real-movement systems is known to reduce sensory conflict, which is a primary contributor to VR-induced nausea. All of these qualities also apply to the VR Enhanced condition, as we were not testing the effects of environmental variations.

The VR Basic environment provided these features:

- a main menu to enable selecting collection, binary, and function
- the ability (via the main menu) to spawn text slates to show program file stats, program strings, function disassembly, or function decompilation
- notepad (associated affordance from Chapter 2: note taking)
- the ability to spawn function call graphs (associated affordance from Chapter 2: spatial semantics)

These features were identified in our development iterations to be minimally viable for performing rudimentary binary analysis.

Spawned objects appeared in fixed positions along a semicircular arc around the center of the working area following a spatial layout designed to emulate a horizontally scrolling interface (similar to the web condition). Users could rotate their heads to view different slates but could not reposition or organize them (the objects were legible from anywhere in the user's working area). This constraint eliminated spatial manipulation affordances such as custom layout or grouping. Ultimately, the VR Basic condition represented a minimalist VR interface with feature parity to the Web interface, isolating the effect of immersive display alone without additional spatial capabilities.

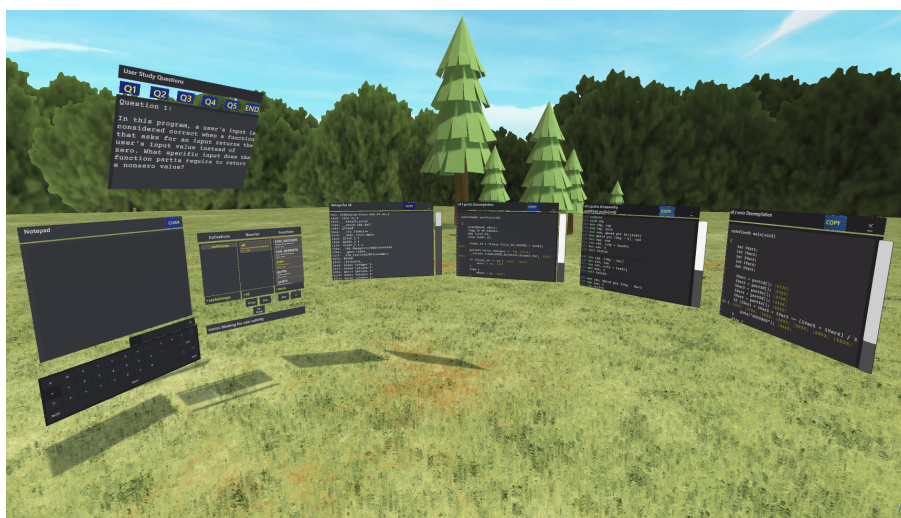


Figure 4.3: VR Basic representative screenshot showing, left-to-right, the notepad, question control, main menu, and four text slates in the semi-circular layout.

VR Enhanced Condition

The VR Enhanced environment built on the Basic configuration and introduced a set of spatial interaction affordances intended to support cognitive offloading, organization, and memory. The additional features included:

- the ability to arbitrarily move and place any object in the environment, including slates, graphs, the main menu, and the notepad, enabling personalized layout and grouping of information (associated affordances from Chapter 2: persistence (spatial memory) and user organization)

- control flow graphs of basic blocks within a function were enabled (associated affordance from Chapter 2: spatial semantics)
- navigation via the function call graph to directly open linked visualizations including disassembly and decompilation slates and control flow graphs (CFGs) (associated affordance from Chapter 2: incremental formalism)
- nodes in the function call graph were visually marked with Capa findings (automated feature detections), helping users identify semantically relevant functions (associated affordance from Chapter 2: signalling)

Together, these enhancements introduced support for spatial memory (by allowing persistent placements), user-driven organization (through arbitrary layout), and incremental formalism (via layered graph exploration). This condition was designed to test whether such affordances support better analytic reasoning or reduce cognitive effort in complex binary comprehension tasks.



Figure 4.4: VR Enhanced representative screenshot showing the default components (notepad, question control, main menu) and additional items spawned and arranged by the user: four text slates, a function call graph, and a control flow graph.

4.1.3 Participants

We recruited participants from the undergraduate and graduate populations within Auburn University's engineering programs. The study targeted, but was not restricted to, students

with foundational knowledge of compiler internals, program analysis, or reverse engineering—topics commonly covered in upper-division or graduate-level coursework. We recruited these individuals via email distribution lists for engineering courses and student organizations, faculty announcements, and peer networks.

We employed a between-subjects design so that each session would require less than one hour (to minimize the impact on busy students) and so that we would only need to design one RE challenge. In contrast, a within-subjects design would have required each participant to spend over 2 hours total in one or more sessions and we would need three unique but equivalent (for assessment purposes) RE challenges to minimize learning effects between conditions, which we determined was not practical.

To determine the number of participants needed, we conducted a power analysis. In the proposal for this work, we ambitiously aimed to detect just-below-medium-sized effects, with estimated Cohen's $d = 0.44$ [104], where $d = 0.5$ is conventionally considered “medium.” Using the `FTestAnovaPower.solve_power` method in the `Statsmodels` library [105] in Python, assuming $d = 0.44$ and commonly-accepted values of significance level $\alpha = 0.05$ and statistical power $1 - \beta = 0.8$, a minimum of 18 participants per condition would be required to detect such effects across three groups using a test such as one-way ANOVA. This sample size aligns with participant counts used in prior empirical VR studies involving cognitive tasks among computing and engineering populations, for example, Romano et al. [89] and Hoff et al. [91] employed 41 and 54 participants, respectively, in a three-condition design for evaluating source code and architecture comprehension in VR.

We repeatedly advertised the study to approximately 1000 students over a three-month period and did not exceed 33 participants, so we elected to proceed with our analysis with that total sample of 33 participants, equally divided among the three conditions. While this sample size limits statistical power for small effects, it is sufficient to observe just-above-medium-sized effects: power analysis using the same method as above indicates that with 11 participants per condition in a between-subjects test such as ANOVA, and with the same commonly-accepted values of significance level $\alpha = 0.05$ and statistical power $1 - \beta = 0.8$, effects of Cohen's $d = 0.57$ (medium) can be detected. We decided this level of sensitivity, still considered “medium,”

is consistent with Cohen’s assertion that medium effect sizes are commonly observed in behavioral science [104] and was sufficiently close to proceed with analysis.

4.1.4 RE Challenge

To evaluate the effectiveness of the system configurations in supporting binary analysis, we designed a reverse engineering (RE) challenge in which participants were asked to analyze a custom-compiled binary program and answer a set of scaffolded questions. The task was designed to represent a simplified, yet realistic, program comprehension scenario encountered in binary RE while remaining accessible to upper-level undergraduate and graduate students. It only relied on the ability to read basic C or ASM code and perform simple math operations.

Participants employed static analysis only. All activities were to be based on the provided visual artifacts: disassembly and decompilation listings, function call graphs, and control flow graphs (in the enhanced VR condition only). This constraint ensured consistency across conditions and focused the evaluation on participants’ ability to synthesize structure and logic from static code representations.

The RE challenge itself was capped at 30 minutes in order to minimize possible negative effects of extended VR usage in a population with little-to-no exposure to VR, respect the time of participants, and facilitate execution of a sufficient number of sessions. Adding time for consent, screening, familiarization, and a post-task survey, total session durations typically lasted 45 to 60 minutes. Participants were free to stop at any time, and all collected data was based on the analysis they completed within the allowed time.

Challenge Binary Program

The binary under analysis was a small C program compiled for x86-64 Linux using gcc with -O0 optimization (none). The full source code is provided in Appendix A. The program’s structure was specifically designed to require users to trace multiple chains of function calls and to engage in simple arithmetic reasoning. It consisted of five parallel chains of functions, each

originating in a user-input function (e.g., part1a, part2a, etc.), and culminating in a shared central function that evaluated whether the values they returned met a set of conditions to produce a success or failure outcome. Each chain performed a sequence of deterministic mathematical transformations on an input integer, and the final function combined their outputs to determine correctness. The modular, compositional structure of the binary encouraged users to analyze subcomponents before understanding the full logic. The design was intended to stimulate spatial reasoning and visual comparison, especially in the VR conditions, where participants could spatially organize multiple code representations and follow multi-step logic flows. A call graph of the program is shown in Figure 4.5.

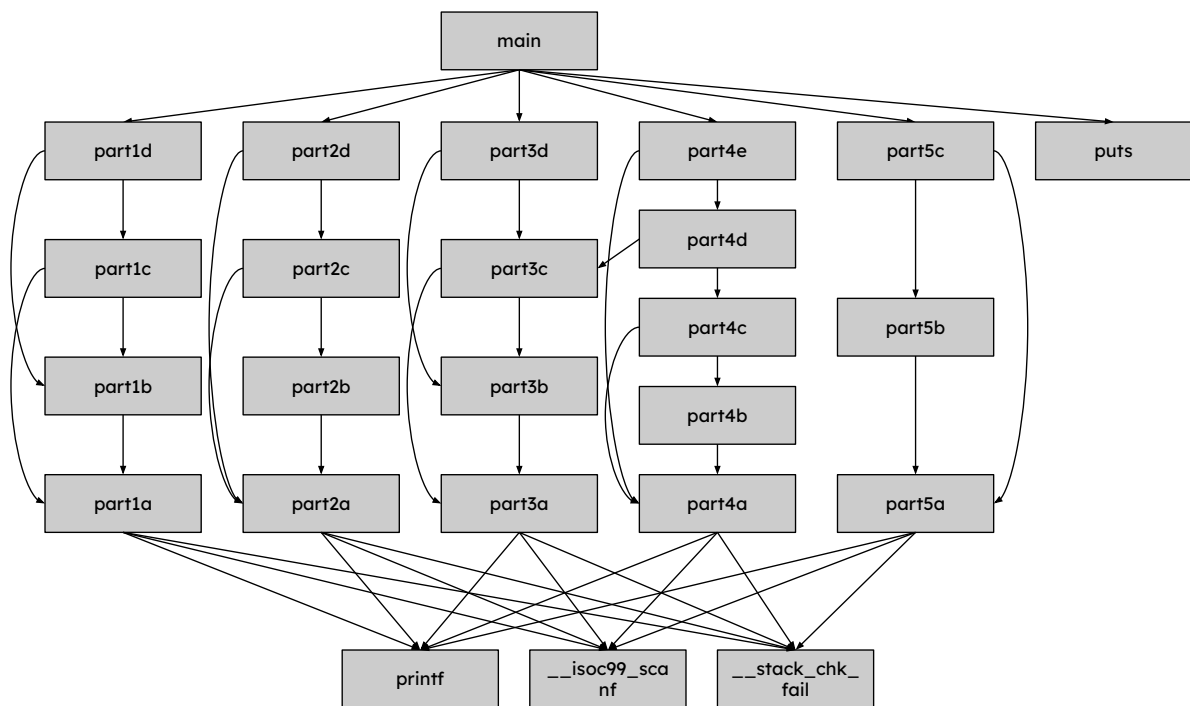


Figure 4.5: Call graph for the “v8” binary used for the reverse engineering challenge

Although simplified, the program mirrors key patterns found in real-world reverse engineering scenarios. Its structure and logic were inspired by the following application domains:

- **Backdoor Detection and Activation:** The requirement to find specific input values to trigger a success condition mimics hidden logic found in malware and backdoors. We presented the program as a “backdoor exploit” to the participants.

- **Obfuscated Binary Analysis:** The program’s use of layered function chains and irrelevant transformations resembles techniques used to obscure logic.
- **License Key Validation:** The stepwise evaluation of user inputs simulates commercial license key checks.
- **Multi-Condition Malware Activation:** The final success function evaluates multiple intermediate values, a pattern seen in stealthy threats and embedded firmware.
- **Capture-The-Flag (CTF) “Crackme”-Style Challenges:** The binary was designed to emulate the pedagogical structure of reverse engineering exercises used in CTF competitions.
- **Embedded System Control Flows:** The deterministic input-response logic reflects patterns seen in hardware-level boot sequences and configuration validators.
- **Security Debugging Workflows:** The necessity to trace values through chains of functions encourages the development of mental models for vulnerability analysis.

By emulating these common real-world structures while retaining pedagogical clarity, the challenge served as a valid surrogate task for evaluating spatial affordance use in immersive reverse engineering. Each question was designed to encourage exploration of different system features (e.g., decompilation, call graphs), require multi-step logical inference across function chains, and stimulate use of spatial affordances in the VR environments.

Challenge Questions

To guide participants through the reverse engineering challenge and support analysis of their problem-solving strategies, we designed a set of five scaffolded questions. These questions increased in difficulty and were intended to gradually expose participants to more complex reasoning paths within the binary. In the context of binary analysis, these questions functioned as *surrogate hypotheses*, standing in for questions that an analysts would ask themselves in the abductive iteration process of exploring a binary while remaining solvable within a 30-minute

task window. The set of questions was designed to have high sensitivity in that we did not anticipate that most participants would answered then all correctly or all incorrectly. The questions and rationale are as follows:

- **Question 1:** “In this program, a user’s input is considered correct when a function that asks for an input returns the user’s input value instead of zero. What specific input does the function part1a require to return a nonzero value?” (Answer: 5)

Rationale: Q1 introduces participants to basic use of the system. It prompts them to locate a specific function and read its decompiled logic. The goal is to familiarize users with environment controls while generating a straightforward success.

- **Question 2:** “What final value is returned by part1d when the input to part1a is correct?” (Answer: 9)

Rationale: Q2 builds on Q1 by requiring the participant to trace the output of the input function (part1a) through a series of three additional functions, each performing simple arithmetic. This tests the ability to follow a static data flow.

- **Question 3:** “What final value is returned by part3d when the input to part3a is correct?” (Answer: 6)

Rationale: Q3 is structurally similar to Q2 but requires the participant to independently find a new starting point (part3a) and repeat the trace. This assesses transfer of strategy and navigation fluency.

- **Question 4:** “What final value is returned by part4e when the required inputs are correct?” (Answer: 40)

Rationale: Q4 increases difficulty by involving five chained functions and requiring either memory or recomputation of prior results from the part3 chain. This tests integration and working memory, which are targets of spatial offloading benefits in VR.

- **Question 5:** “How many times is the function part1a called if the part1d function fully executes (you examined this chain of functions in question 2)?” (Answer: 4)

Rationale: Q5 departs from arithmetic logic and probes structural comprehension. Participants must inspect the call graph or function bodies to determine that a specific function is invoked multiple times by its successors, requiring deeper reasoning.

The questions were presented both visually (via a question display user interface element in the environment) and verbally by the facilitator. Participants responded aloud and their answers were recorded by the researcher. If an answer was incorrect, the correct answer was provided to allow participants to continue to the next question with an accurate understanding of the code's logic.

4.1.5 Data Collection

To comprehensively assess participant performance, perception, and interaction behavior across conditions, we employed a multi-modal data collection approach combining subjective survey instruments, system-generated logs, and direct observation. Data collection focused on capturing both quantitative metrics and qualitative usage patterns relevant to the use of spatial affordances in immersive environments. These measures are our dependent variables while selection of condition is the independent variable.

We captured extensive interaction logs and spatial telemetry:

- **Task Performance Metrics:** For each participant, investigators recorded accuracy (number of correct answers out of five questions) manually because participants answered the challenge questions verbally. The system recorded completion time (duration, in seconds, between the times that the user received each question). These values were used to test Hypothesis 1.
- **Interaction Events:** Automatically recorded user actions included the creation, movement, and deletion of visual objects (slates, graphs, notepad, etc.). These logs enabled analysis of behavioral strategies and usage frequency of specific features.
- **Observations:** Investigators manually recorded a log of notable interactions and quotations as participants performed the RE challenge.

- **Spatial Telemetry (in VR Conditions):** In the VR Basic and VR Enhanced environments, the system logged positional and rotational data for the user's head and hands (collected every 0.15 seconds, or 6.67 frames per second), as well as the spatial coordinates of all spawned objects (collected every 0.5 seconds, or 2 frames per second). This data was used to compute metrics, e.g., total workspace area or volume used in VR, which is critical for testing Hypothesis 4, which posits a relationship between spatial use and task performance.

After completing a session, participants answered a series of validated and custom-designed surveys:

- **NASA Task Load Index (NASA-TLX):** This survey [84] measured perceived cognitive workload across six dimensions: mental demand, physical demand, temporal demand, performance, effort, and frustration. These responses were used to assess Hypothesis 2.
- **System Usability Scale (SUS):** Adapted from Brooke [106], this ten-item instrument assessed perceived usability of the system in each condition.
- **Presence Questionnaire (PQ):** Applicable items from Witmer and Singer's Presence Questionnaire [107] evaluated subjective immersion and the extent to which participants felt "present" in the VR environment (collected only for the VR conditions).
- **Affordance Usefulness Ratings:** A custom Likert-scale questionnaire asked participants in the Enhanced VR condition (only) to rate the perceived usefulness of individual affordances (e.g., object placement, graph linking, signaling) with respect to task understanding and information recall. These responses were used to assess Hypothesis 3 and to identify features perceived as most valuable by users.

All survey data was collected electronically using Qualtrics, an IRB-approved and Auburn-licensed platform. Survey contents are available in the IRB package in Appendix C. Table 4.1 summarizes the data we collected.

Instrument	Num. Variables	Value characterization	Condition(s) in which collected
Pre-survey	6 + mean	Self-rating: 1 (lowest) to 5 (highest)	All
Accuracy	5 + mean	RE question response: 0 (incorrect) or 1 (correct)	All
Completion Time	5 + mean	Time in seconds to answer RE question	All
TLX	6 + mean	Self-rating: 0 (lowest load) to 100 (highest)	All
SUS	10 + mean	Rating: 1 (strongly disagree) to 5 (strongly agree)	All
Presence	23 + mean	Rating: 1 (not at all) to 7 (completely)	VR only
Affordance Ratings	6	Rating: 1 (least valuable) to 5 (most valuable)	VR Enhanced only
Interaction Events	N/A	Discrete irregular events	All
Spatial Telemetry	N/A	Discrete regular events	VR only
Investigator Observations	N/A	Notes	All

Table 4.1: Summary of the data we collected.

This multi-source data collection strategy enabled triangulation of results across performance, perception (surveys), and behavior (logs), providing a robust foundation for analysis of the effects of immersive spatial affordances on binary RE.

4.1.6 Procedure

Each participant completed a single 45-to-60-minute session, which included informed consent, orientation to the assigned condition, execution of the reverse engineering challenge, and post-task survey completion. The procedure was standardized and documented in a detailed User Study Playbook, included in Appendix B, to ensure consistency across sessions and experimenters. Sessions were conducted in a controlled lab environment in the Shelby Center for Engineering Technology at Auburn University. Two investigators traded facilitation duties after co-leading initial sessions to ensure calibration of instruction and observation techniques.

Each session flow proceeded as follows:

- **Pre-Session Preparation:** Researchers verified that the appropriate software environment (desktop or VR) was running and correctly configured. For VR sessions, the configuration file (config.ini) was checked to ensure that either the Basic or Enhanced feature set was active. For the desktop condition, the Flask/React web interface was launched with the appropriate binary ("v8") preloaded for analysis.
- **Consent and Screening:** Participants signed two consent forms: One via the pre-survey platform (Qualtrics), covering demographic data and contact information, and another

on paper upon arrival, covering participation in the interactive session, data logging, and follow-up analysis. Any participant with a reported or suspected contraindication to VR use, as listed in a screening tool, was reassigned or excluded to ensure safety.

- **Condition Assignment:** Participants were assigned to one of the three study conditions: Desktop (Web), VR Basic, or VR Enhanced. The assigned condition determined the interaction device (monitor/mouse vs. VR headset/controllers) and available affordances.
- **Environment Orientation:** Participants viewed a short introduction video tailored to their assigned condition. This explained how to navigate the interface and use available features (e.g., spawning slates, accessing function views). In VR conditions, participants were fitted with the headset and hand controllers, and briefly acclimated to the immersive environment. The facilitator ensured that participants could interact with basic elements before starting the challenge.
- **Reverse Engineering Challenge:** Participants initiated the task by clicking the “Q1” button to reveal the first question when prompted by the investigator. Each of the five scaffolded questions was delivered via in-environment prompts. Participants responded aloud. The facilitator recorded the correctness of each response and optionally provided corrective feedback for incorrect answers to allow continued progression. During the task, system logs captured all user interactions, and researchers took note of any noteworthy behaviors, strategies, or verbal comments.
- **Post-Task Survey:** After the challenge, participants completed the post-task survey on a lab workstation. Survey content varied slightly based on condition (e.g., affordance usefulness questions were only presented in the Enhanced VR condition).
- **Session Wrap-Up:** The investigator confirmed the participant was not suffering any ill-effects from participation (and took appropriate actions if they were), and then ensured that all data logs were successfully saved and uploaded to the secure project folder on the cloud storage system. Participants were thanked for their time and provided with a

\$20 electronic gift card via a batch processing system within one week of completion.

Between VR sessions, all equipment was cleaned and sanitized according to lab protocol.

This consistent and carefully managed procedure ensured high data quality, participant safety, and experimental rigor across all three conditions.

4.1.7 Approval

This study was approved by the Auburn Institutional Review Board. The stamped IRB package is included in Appendix C.

4.1.8 Threats to Validity

As with any experimental study, this investigation is subject to a number of threats to validity. Careful design choices and procedural safeguards were implemented to mitigate these threats, but they remain important considerations when interpreting results. We identified threats in four common categories.

Internal validity concerns the extent to which observed effects can be attributed to a causal relationship between variables rather factors such as confounding variables or experimental manipulation. While the between-subjects design minimized learning effects as each participant experienced only one condition, it also increased susceptibility to individual differences. Participants were assigned to conditions without regard to prior experience or task familiarity, helping to evenly distribute those potentially confounding variables. Pre-task instructions and tutorial videos were standardized across sessions to control for variation in investigator explanation. The task design was intentionally constrained (e.g., no binary execution allowed, static analysis only) to limit alternate problem-solving strategies that could confound comparison between conditions. While participant experience levels were not strictly matched across groups, pre-survey data indicated comparable distributions of related skills across conditions (as discussed in Section 4.1.3, reducing the likelihood of major selection bias.

External validity refers to the generalizability of the findings to broader populations and settings. Our threats result from practical considerations. The participant pool, primarily upper-level undergraduates and graduate students in technical disciplines, limits generalizability to the target audience, industry professionals, of this broader line of research; however, as experts are geographically dispersed and the sessions require in-person participation, we drew from the next-best population. As a result, the RE task was synthetic and shorter in duration than real-world reverse engineering problems, which often span hours or days and involve deeper system knowledge; however, the design mimicked common reasoning structures found in real scenarios (e.g., multi-step function chains, obfuscated logic, control flow inference). Finally, we built the immersive portion of our system from scratch with limited resources; while we made our design decisions in implementing the VR experience (e.g., environment, interaction, and visualization choices) following our iterative DT/CSE process that incorporated foundational research and the voice of binary RE practitioners, we implemented only a limited subset of affordances in VR and our findings may not generalize to a hypothetical more fully-featured VR system supporting binary RE.

Construct validity assesses whether the measures used accurately capture the intended theoretical constructs. Performance was measured as correctness and time to solve a constrained set of scaffolded RE questions; however, while valid for assessing procedural comprehension, this may not fully reflect deeper conceptual understanding, mental models, or exploratory strategies. Cognitive load was assessed using the NASA-TLX, a validated and widely used subjective instrument; however, it may not capture transient or localized workload fluctuations during spatial interaction. The usefulness of affordances was assessed through post-task Likert ratings; however, while helpful in gauging perceived value, these ratings are inherently subjective and may not correlate directly with behavioral use. Spatial usage was quantified using artifact placement telemetry (area in VR Basic, volume in VR Enhanced); while these metrics provide a meaningful proxy for externalized cognitive organization, though they do not account for quality or intent of layout. Where possible, we used both self-reported (subjective) data and observable actions (behavioral data) to support construct validity through converging evidence.

Statistical conclusion validity concerns whether the statistical inferences drawn from the data support the conclusion. The sample size ($n = 11$ per condition) was below the original power analysis target ($n = 18$ per group), which moved our detection power from just-below-medium-sized effects to just-above-medium-sized effects. The sample size still supports medium sensitivity and is common in similar research as discussed in Section 4.1.3. We conducted analyses using appropriate statistical tests (e.g., ANOVA for between-group comparisons, correlation for spatial usage vs. performance), and reviewed assumptions such as normality and variance homogeneity prior to interpretation. In the presence of non-significant findings (particularly in performance and cognitive load comparisons), we avoided overinterpreting null results and emphasized effect sizes and observed trends. Despite these limitations, the study's design, instrumentation, and analysis methods provide a solid basis for preliminary conclusions and inform future work with clearer design constraints and hypotheses.

4.2 Data and Analysis

This section presents the data collected from the study and then its analysis, including both quantitative metrics and qualitative observations.

4.2.1 Data Overview

Here we provide a descriptive overview of the data collected across the three experimental conditions (mean, standard deviation, median, range, number of observations) without interpretation. The data are presented in tables and charts and for presentation reasons, many of the column headings are abbreviated. In the case of survey data, these columns are presented in the same order the questions were asked, as listed in the survey instruments in the IRB in Appendix C. Taken together, these data offer a multidimensional perspective on how immersive environments impact both the objective performance and subjective experience of binary RE practitioners.

Participant Backgrounds

Interested individuals completed a pre-survey that collected information about their technical background. The full text of the pre-survey is in the IRB package in Appendix C. Of the 33 participants, 26 were enrolled in computing-related degree programs: 22 in Computer Science or Software Engineering, and 4 in Computer Engineering. The remaining 7 participants included students from Electrical Engineering ($n = 3$), Cybersecurity Engineering ($n = 1$), Aerospace Engineering ($n = 1$), Mechanical Engineering ($n = 1$), and Pre-Pharmacy ($n = 1$). This reflects a participant pool predominantly drawn from computing disciplines, but with some interdisciplinary representation.

The pre-survey included self-assessment of experience in six technical areas, rated on a 1–5 scale (1 = None, 5 = Expert). Table 4.2 provides a statistical overview of the responses and Figures 4.6 and 4.7 visualize them. The group reported moderate experience in software development ($M = 3.24$, $SD = 1.06$) and computer architecture ($M = 2.88$, $SD = 1.08$). Experience with compiler construction ($M = 2.16$, $SD = 0.88$) and binary reverse engineering ($M = 2.12$, $SD = 1.08$) was lower, consistent with expectations for a student sample. VR exposure was also limited: participants reported low-to-moderate experience as VR users ($M = 2.33$, $SD = 1.27$) and minimal experience as VR designers or developers ($M = 1.45$, $SD = 0.97$). These distributions suggest that participants as a group approached the task with relevant but non-expert technical knowledge, making them an appropriate population for performing the RE challenge designed with these expectations in mind.

Performance Metrics

Performance was assessed based on both accuracy (correctness) and completion time for each RE challenge question. Across conditions, the average number of correct answers ranged from 0.65 to 0.67 (on a 0–1 scale), with standard deviations around 0.27–0.29. There was no evident difference in accuracy between Web and VR conditions. Table 4.3 provides a statistical overview of accuracy for each condition, while Figure 4.8 visualizes the summary and

Condition	Statistic	SW Dev	Comp Arch	Compiler Const	Binary RE	VR User	VR Dev	Average
VR Enhanced	mean	3.18	2.64	2.1	2.27	2.64	1.27	2.34
	std	0.98	1.03	0.99	1.1	1.5	0.65	0.74
	median	3.0	2.0	2.0	2.0	2.0	1.0	2.0
	min	2.0	1.0	1.0	1.0	1.0	1.0	1.33
	max	5.0	4.0	4.0	4.0	5.0	3.0	3.5
	count	11.0	11.0	10.0	11.0	11.0	11.0	11.0
VR Basic	mean	3.36	2.64	2.18	2.0	2.64	1.91	2.45
	std	1.21	1.12	0.75	0.77	1.36	1.45	0.91
	median	3.0	3.0	2.0	2.0	3.0	1.0	2.5
	min	1.0	1.0	1.0	1.0	1.0	1.0	1.17
	max	5.0	5.0	3.0	3.0	5.0	5.0	4.33
	count	11.0	11.0	11.0	11.0	11.0	11.0	11.0
Web	mean	3.18	3.36	2.18	2.09	1.73	1.18	2.29
	std	1.08	1.03	0.98	1.38	0.65	0.4	0.69
	median	3.0	3.0	2.0	1.0	2.0	1.0	2.17
	min	2.0	2.0	1.0	1.0	1.0	1.0	1.5
	max	5.0	5.0	4.0	4.0	3.0	2.0	3.83
	count	11.0	11.0	11.0	11.0	11.0	11.0	11.0

Table 4.2: Statistical overview of self-reported skill levels in relevant areas (on scale 1-5).

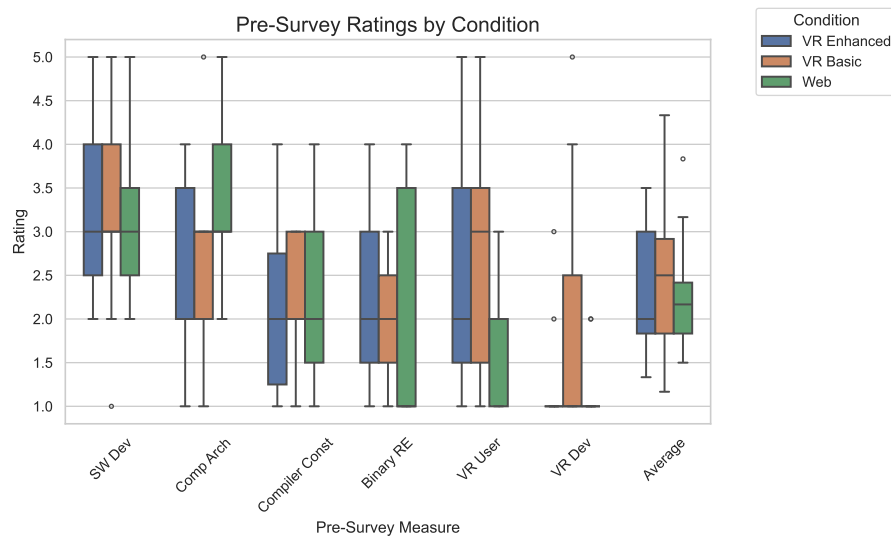


Figure 4.6: Pre-survey ratings box plot.

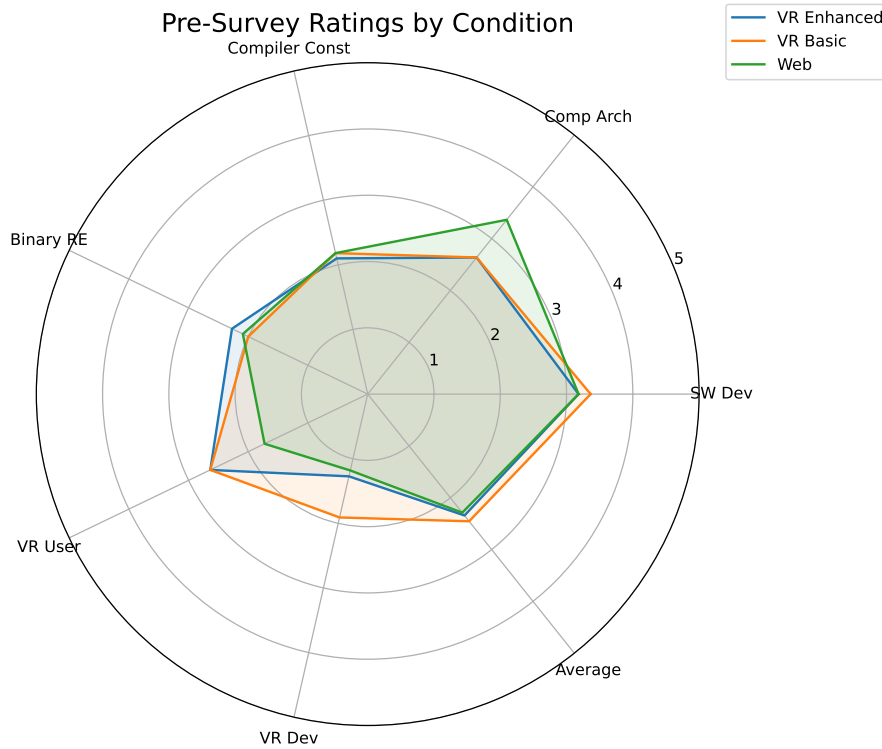


Figure 4.7: Pre-survey ratings radar plot.

Figure 4.9 shows accuracy for each question. Average task durations were highest in VR Enhanced (273.63s) and VR Basic (267.20s), and shortest in the Web condition (221.11s), with notable variance within all groups. Table 4.4 provides a statistical overview of duration for each condition, while Figure 4.10 visualizes the summary and Figure 4.11 shows duration for each question.

Cognitive Load (NASA-TLX) scores

Cognitive load was measured using the NASA-TLX scale administered as a post-task online survey. Table 4.5 provides a statistical overview of each subscale for each condition, while Figures 4.12 and and Figure 4.13 shows visualize the data. Average cognitive load ratings (0–100 scale) were VR Enhanced, 31.47; VR Basic, 32.41; and Web, 40.73. Participants in the Web condition reported higher mental demand and effort than those in either VR condition.

Condition	Statistic	Q1 Correct	Q2 Correct	Q3 Correct	Q4 Correct	Q5 Correct	Avg Correct
VR Enhanced	mean	0.91	0.73	0.82	0.45	0.45	0.67
	std	0.3	0.47	0.4	0.52	0.52	0.29
	median	1.0	1.0	1.0	0.0	0.0	0.8
	min	0.0	0.0	0.0	0.0	0.0	0.2
	max	1.0	1.0	1.0	1.0	1.0	1.0
	count	11.0	11.0	11.0	11.0	11.0	11.0
VR Basic	mean	1.0	0.55	0.91	0.45	0.36	0.65
	std	0.0	0.52	0.3	0.52	0.5	0.27
	median	1.0	1.0	1.0	0.0	0.0	0.6
	min	1.0	0.0	0.0	0.0	0.0	0.2
	max	1.0	1.0	1.0	1.0	1.0	1.0
	count	11.0	11.0	11.0	11.0	11.0	11.0
Web	mean	0.91	0.82	0.82	0.55	0.27	0.67
	std	0.3	0.4	0.4	0.52	0.47	0.27
	median	1.0	1.0	1.0	1.0	0.0	0.8
	min	0.0	0.0	0.0	0.0	0.0	0.2
	max	1.0	1.0	1.0	1.0	1.0	1.0
	count	11.0	11.0	11.0	11.0	11.0	11.0

Table 4.3: Statistical overview of correctness of RE Challenge responses.

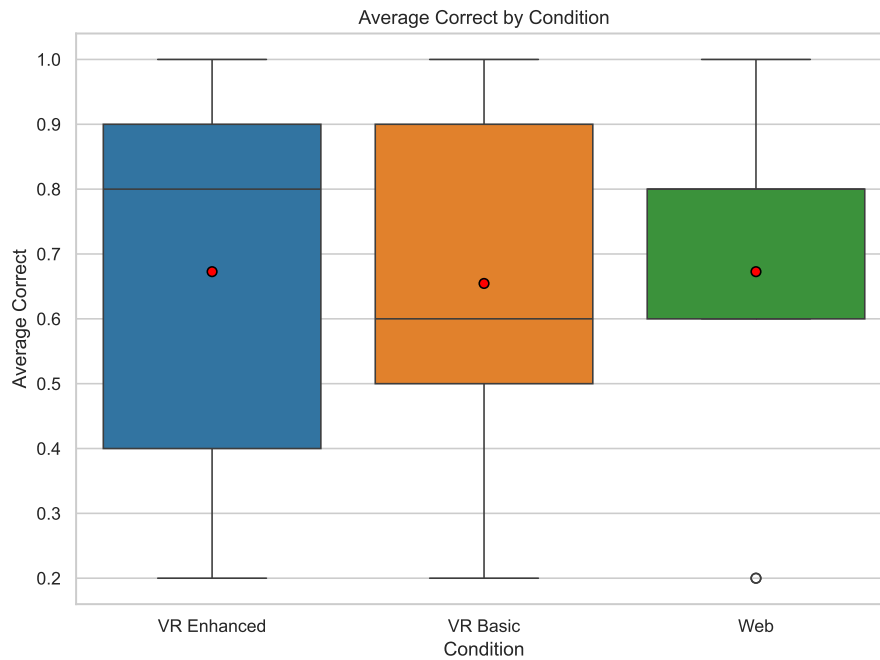


Figure 4.8: RE Challenge Question Correctness box plot.

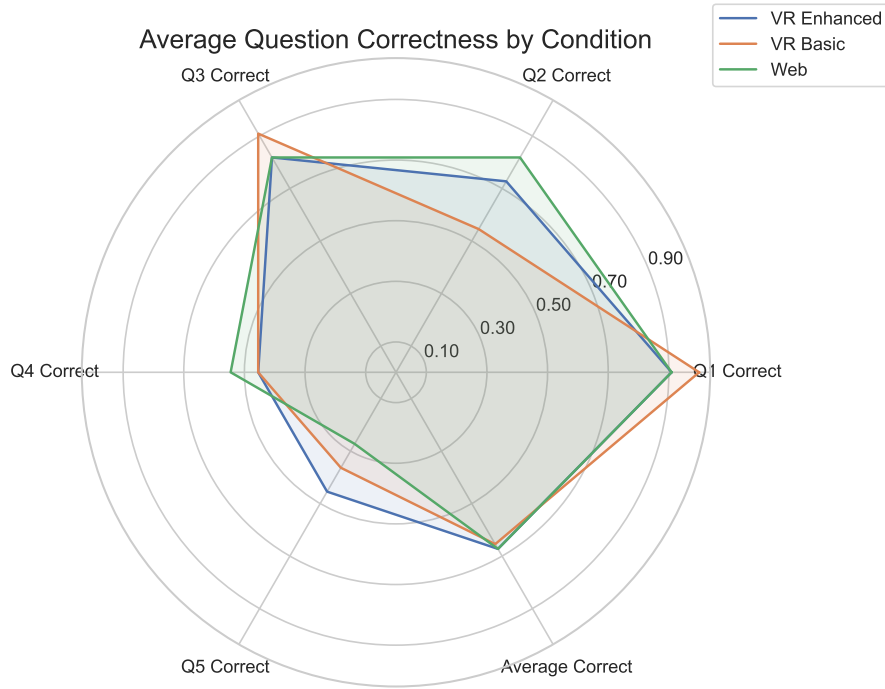


Figure 4.9: RE Challenge Question Correctness radar plot.

Condition	Statistic	Q1 Duration	Q2 Duration	Q3 Duration	Q4 Duration	Q5 Duration	Avg Duration
VR Enhanced	mean	278.31	240.49	214.64	478.62	156.09	273.63
	std	241.33	93.16	65.73	178.05	67.97	88.56
	median	209.0	213.0	196.45	454.96	130.56	263.71
	min	61.17	110.72	101.0	161.27	77.6	118.52
	max	853.61	409.3	297.25	783.9	296.97	430.35
	count	11.0	11.0	11.0	11.0	11.0	11.0
VR Basic	mean	241.57	244.17	234.66	454.49	161.13	267.2
	std	163.39	117.78	106.58	125.81	44.35	76.77
	median	162.63	240.0	239.67	407.08	144.02	264.0
	min	84.54	116.1	91.94	273.59	120.0	159.96
	max	620.82	547.6	480.0	674.57	255.42	452.49
	count	11.0	11.0	11.0	11.0	11.0	11.0
Web	mean	241.76	168.93	203.34	363.01	128.5	221.11
	std	94.68	63.0	56.97	164.94	33.39	52.56
	median	242.53	130.34	212.81	364.44	128.76	211.24
	min	85.53	98.85	134.16	152.83	75.55	128.45
	max	400.07	290.07	278.88	601.26	202.72	296.56
	count	11.0	11.0	11.0	11.0	11.0	11.0

Table 4.4: Statistical overview of duration of RE Challenge responses (seconds).

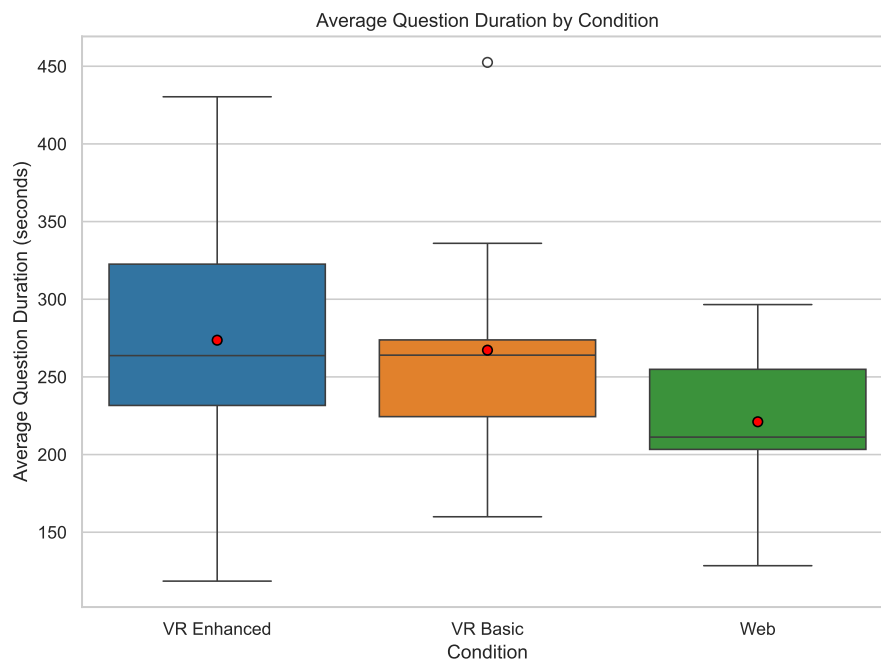


Figure 4.10: RE Challenge Question Duration box plot.

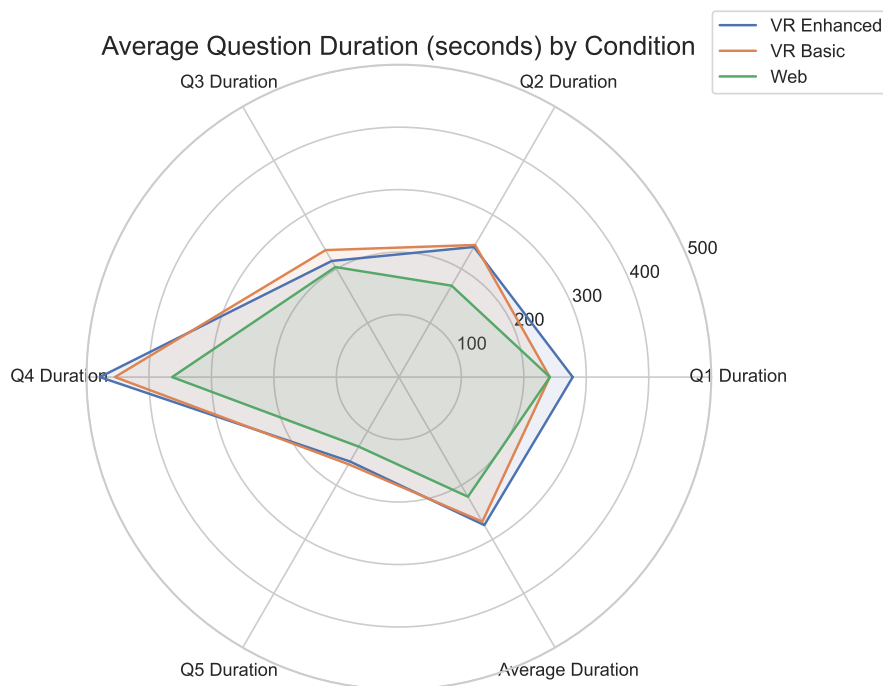


Figure 4.11: RE Challenge Question Duration radar plot.

Condition	Statistic	Mental Demand	Phys Demand	Temp Demand	Perf	Effort	Frust	Average
VR Enhanced	mean	48.18	25.18	19.09	30.36	41.09	23.7	31.47
	std	25.57	22.96	16.9	26.2	23.24	17.52	13.98
	median	50.0	15.0	20.0	20.0	50.0	22.5	37.0
	min	7.0	0.0	0.0	0.0	7.0	0.0	3.17
	max	80.0	60.0	50.0	80.0	80.0	50.0	46.67
	count	11.0	11.0	11.0	11.0	11.0	10.0	11.0
VR Basic	mean	39.91	21.0	18.55	38.55	47.18	29.27	32.41
	std	15.47	16.22	14.01	32.46	18.59	32.5	10.28
	median	40.0	19.0	14.0	33.0	45.0	11.0	29.33
	min	18.0	5.0	0.0	0.0	20.0	1.0	19.5
	max	65.0	52.0	42.0	100.0	75.0	92.0	51.67
	count	11.0	11.0	11.0	11.0	11.0	11.0	11.0
Web	mean	48.0	19.43	23.55	65.18	51.36	29.0	40.73
	std	20.89	21.88	19.35	26.4	23.55	24.04	8.76
	median	42.0	10.0	20.0	80.0	51.0	20.0	41.0
	min	20.0	2.0	4.0	10.0	20.0	5.0	31.17
	max	86.0	66.0	70.0	90.0	80.0	70.0	57.83
	count	11.0	7.0	11.0	11.0	11.0	11.0	11.0

Table 4.5: Statistical overview of TLX responses (on scale 0-100).

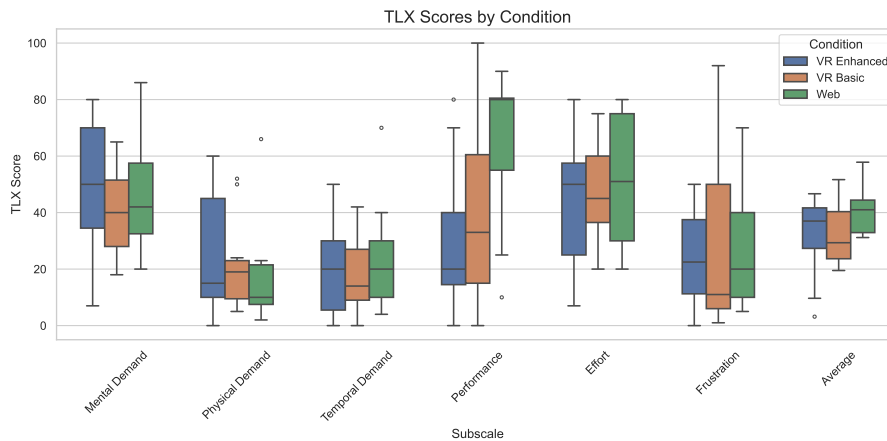


Figure 4.12: TLX box plot.

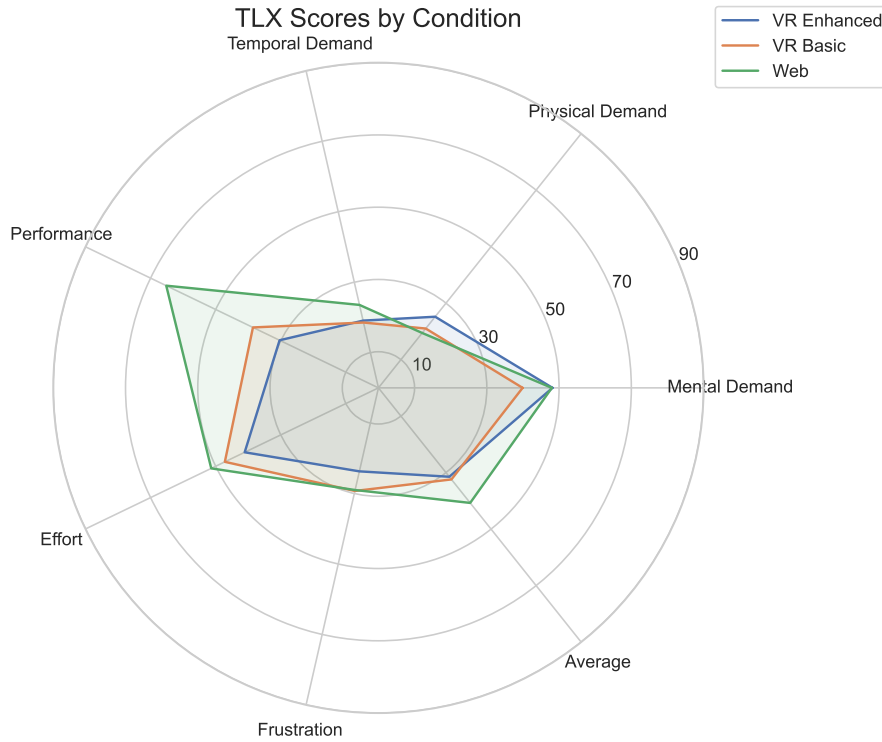


Figure 4.13: TLX radar plot.

System Usability Scale (SUS) scores

Participants completed the SUS survey post-task. As the SUS has positive sentiment questions (higher score is better) and negative sentiment questions (lower score is better), we present the data in two parts. The positive sentiment data summary is in Table 4.6 and visualized in Figure 4.14, while the negative sentiment data summary is in Table 4.7 and visualized in Figure 4.15. VR Basic showed slightly higher negative-sentiment scores; otherwise, the three conditions are rated similarly.

Presence Questionnaire scores

Only participants in the VR (Enhanced and Basic) conditions completed the post-task Presence questionnaire, in which we asked 23 questions that apply to our use case. For presentation purposes, we split the statistical overview of the data into three tables; see Tables 4.8 to 4.10. We used numbers instead of the question text in the column headers for tables and plots to identify each subscale; the key is in Table 4.11. We provide visualizations in Figure 4.16 and Figure 4.17.

Condition	Statistic	Use Freq.	Easy To Use	Well Integ.	Learn Quickly	Confident
VR Enhanced	mean	2.82	4.0	4.09	4.18	3.91
	std	1.17	1.25	0.94	0.87	1.04
	median	3.0	4.5	4.0	4.0	4.0
	min	1.0	2.0	2.0	3.0	2.0
	max	5.0	5.0	5.0	5.0	5.0
	count	11.0	10.0	11.0	11.0	11.0
VR Basic	mean	3.09	4.18	3.55	4.27	4.18
	std	1.14	0.75	1.29	1.01	1.08
	median	3.0	4.0	4.0	5.0	5.0
	min	1.0	3.0	1.0	2.0	2.0
	max	5.0	5.0	5.0	5.0	5.0
	count	11.0	11.0	11.0	11.0	11.0
Web	mean	2.64	4.27	4.0	3.73	4.36
	std	1.43	1.1	1.1	0.9	0.92
	median	2.0	5.0	4.0	4.0	5.0
	min	1.0	2.0	2.0	2.0	2.0
	max	5.0	5.0	5.0	5.0	5.0
	count	11.0	11.0	11.0	11.0	11.0

Table 4.6: Statistical overview of SUS responses to positive-sentiment items (on scale 1-5).

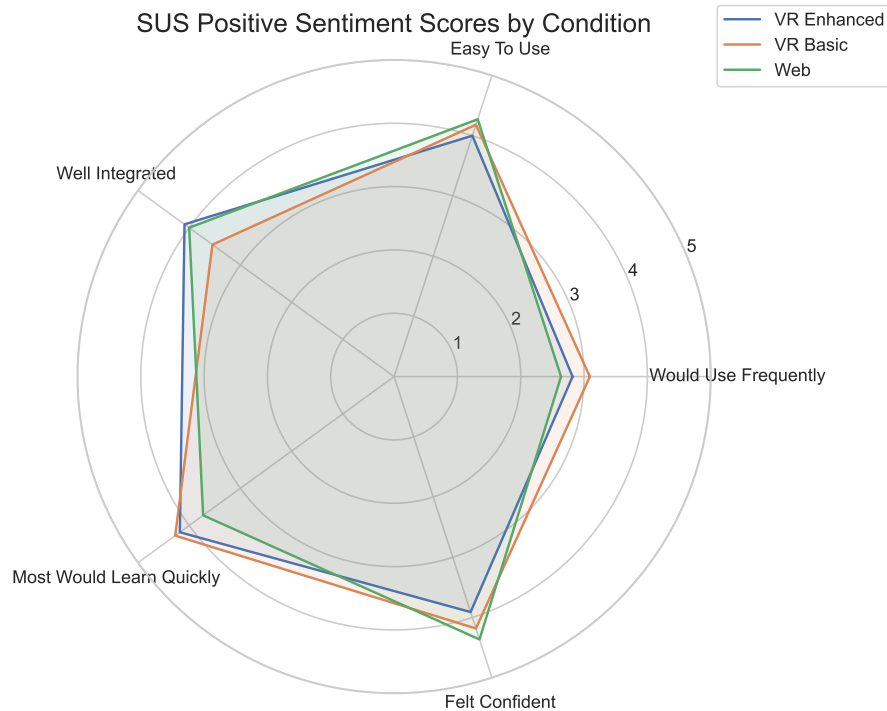


Figure 4.14: SUS positive-sentiment radar plot.

Condition	Statistic	Unnec. Complex	Need Support	Inconsistent	Cumbersome	Learn First
VR Enhanced	mean	2.82	4.0	4.09	4.18	3.91
	std	1.17	1.25	0.94	0.87	1.04
	median	3.0	4.5	4.0	4.0	4.0
	min	1.0	2.0	2.0	3.0	2.0
	max	5.0	5.0	5.0	5.0	5.0
	count	11.0	10.0	11.0	11.0	11.0
VR Basic	mean	3.09	4.18	3.55	4.27	4.18
	std	1.14	0.75	1.29	1.01	1.08
	median	3.0	4.0	4.0	5.0	5.0
	min	1.0	3.0	1.0	2.0	2.0
	max	5.0	5.0	5.0	5.0	5.0
	count	11.0	11.0	11.0	11.0	11.0
Web	mean	2.64	4.27	4.0	3.73	4.36
	std	1.43	1.1	1.1	0.9	0.92
	median	2.0	5.0	4.0	4.0	5.0
	min	1.0	2.0	2.0	2.0	2.0
	max	5.0	5.0	5.0	5.0	5.0
	count	11.0	11.0	11.0	11.0	11.0

Table 4.7: Statistical overview of SUS responses to negative-sentiment items (on scale 1-5).

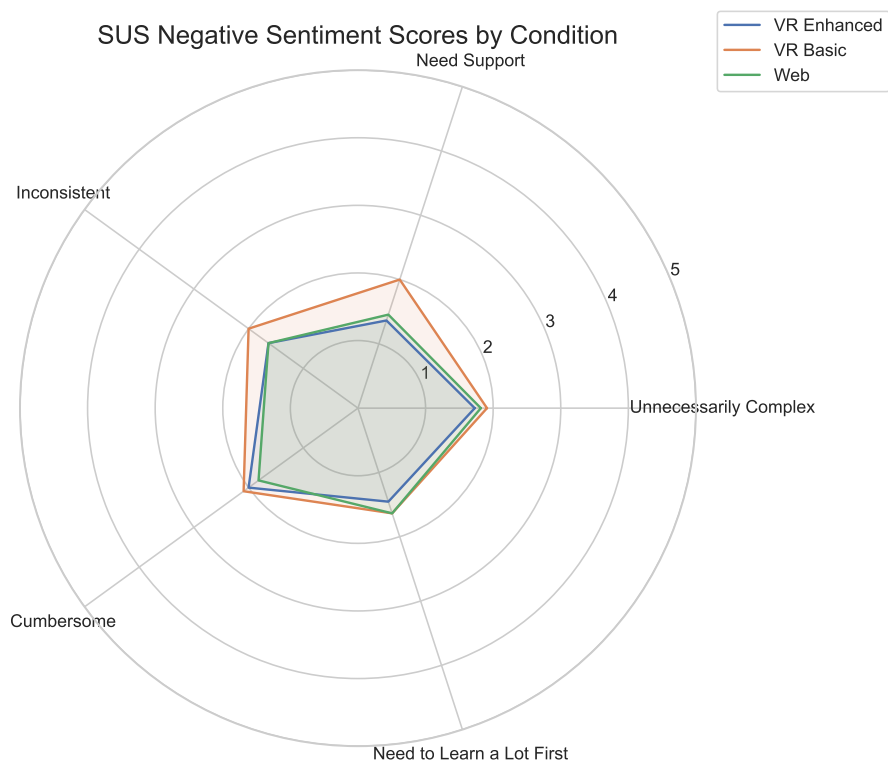


Figure 4.15: SUS negative-sentiment radar plot.

The presence scores averaged 5.16 in VR Enhanced and 5.02 in VR Basic on a 1–7 scale, indicating a strong sense of immersion. While most subscale scores were close between the conditions, VR Enhanced had higher scores in several areas: “How natural was the mechanism which controlled movement through the environment?”, “How compelling was your sense of objects moving through space?”, “How much did your experiences in the virtual environment seem consistent with your real world experiences?”, and “How completely were your senses engaged in this experience?” VR Basic had higher scores in “How completely were you able to actively survey or search the environment using vision?”, “How compelling was your sense of moving around inside the virtual environment?”, and “Was the information provided through different senses in the virtual environment (e.g., vision, hearing, touch) consistent?”

Condition	Statistic	1	2	3	4	5	6	7	8
VR Enhanced	mean	5.82	5.0	5.18	5.18	5.55	6.0	4.45	6.27
	std	0.87	1.48	2.04	2.04	2.11	1.0	1.69	1.01
	median	6.0	5.0	6.0	6.0	6.0	6.0	5.0	7.0
	min	5.0	3.0	1.0	1.0	1.0	4.0	1.0	4.0
	max	7.0	7.0	7.0	7.0	7.0	7.0	7.0	7.0
	count	11.0	11.0	11.0	11.0	11.0	11.0	11.0	11.0
VR Basic	mean	5.91	5.09	4.64	4.73	4.64	4.91	3.73	6.0
	std	1.3	1.51	1.91	1.49	1.86	1.45	1.56	0.89
	median	6.0	6.0	5.0	5.0	5.0	5.0	4.0	6.0
	min	3.0	2.0	1.0	2.0	1.0	3.0	1.0	4.0
	max	7.0	7.0	7.0	7.0	7.0	7.0	6.0	7.0
	count	11.0	11.0	11.0	11.0	11.0	11.0	11.0	11.0

Table 4.8: Statistical overview of responses to first set of Presence items (on scale 1-7).

VR Affordance Ratings

Participants rated the helpfulness of various VR-specific affordances in facilitating task performance in a post-task survey in which they rated their agreement with provided statements in a scale of 1-5. The affordances and their associated survey statements are as follows:

- **Incr[emental] Formalism:** Working in VR, with respect to understanding the problem,

I found the ability to navigate from call graph to control flow graph to code:

Condition	Statistic	9	10	11	12	13	14	15	16
VR Enhanced	mean	5.64	5.27	5.73	5.18	2.91	5.55	5.91	4.0
	std	1.57	1.79	1.62	1.33	1.7	0.93	1.51	2.19
	median	6.0	6.0	6.0	5.0	2.0	6.0	6.0	4.0
	min	2.0	1.0	3.0	4.0	1.0	4.0	2.0	1.0
	max	7.0	7.0	7.0	7.0	6.0	7.0	7.0	7.0
	count	11.0	11.0	11.0	11.0	11.0	11.0	11.0	11.0
VR Basic	mean	6.27	6.0	5.55	5.55	2.91	5.73	6.18	3.45
	std	1.19	0.77	1.97	1.29	1.92	1.19	1.66	1.57
	median	7.0	6.0	6.0	6.0	2.0	6.0	7.0	4.0
	min	3.0	5.0	2.0	3.0	1.0	4.0	2.0	1.0
	max	7.0	7.0	7.0	7.0	6.0	7.0	7.0	6.0
	count	11.0	11.0	11.0	11.0	11.0	11.0	11.0	11.0

Table 4.9: Statistical overview of responses to second set of Presence items (on scale 1-7).

Condition	Statistic	17	18	19	20	21	22	23	Avg
VR Enhanced	mean	3.09	5.27	5.55	4.36	5.82	5.64	5.27	5.16
	std	1.76	1.62	1.29	1.91	1.6	0.81	1.42	0.72
	median	4.0	5.0	6.0	5.0	6.0	6.0	5.0	5.22
	min	1.0	2.0	3.0	1.0	2.0	4.0	3.0	3.39
	max	5.0	7.0	7.0	7.0	7.0	7.0	7.0	6.13
	count	11.0	11.0	11.0	11.0	11.0	11.0	11.0	11.0
VR Basic	mean	3.27	5.0	4.82	4.27	5.64	5.27	5.82	5.02
	std	1.9	1.67	1.25	1.56	1.63	1.35	1.08	0.67
	median	4.0	6.0	4.0	4.0	6.0	5.0	6.0	5.3
	min	1.0	2.0	3.0	2.0	3.0	3.0	3.0	3.87
	max	7.0	7.0	7.0	7.0	7.0	7.0	7.0	5.78
	count	11.0	11.0	11.0	11.0	11.0	11.0	11.0	11.0

Table 4.10: Statistical overview of responses to third set of Presence items (on scale 1-7).

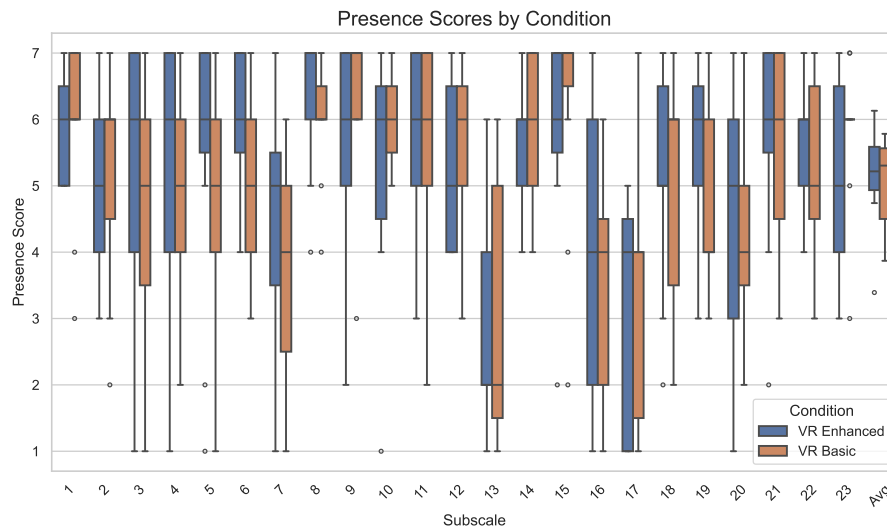


Figure 4.16: Presence box plot.

ID	Presence Question
1	How much were you able to control events?
2	How responsive was the environment to actions that you initiated (or performed)?
3	How natural did your interactions with the environment seem?
4	How much did the visual aspects of the environment involve you?
5	How natural was the mechanism which controlled movement through the environment?
6	How compelling was your sense of objects moving through space?
7	How much did your experiences in the virtual environment seem consistent with your real world experiences?
8	Were you able to anticipate what would happen next in response to the actions that you performed?
9	How completely were you able to actively survey or search the environment using vision?
10	How compelling was your sense of moving around inside the virtual environment?
11	How well could you move or manipulate objects in the virtual environment?
12	How involved were you in the virtual environment experience?
13	How much delay did you experience between your actions and expected outcomes?
14	How quickly did you adjust to the virtual environment experience?
15	How proficient in moving and interacting with the virtual environment did you feel at the end of the experience?
16	How much did the visual display quality interfere or distract you from performing assigned tasks or required activities?
17	How much did the control devices interfere with the performance of assigned tasks or with other activities?
18	How well could you concentrate on the assigned tasks or required activities rather than on the mechanisms used to perform those tasks or activities?
19	How completely were your senses engaged in this experience?
20	How easy was it to identify objects through physical interaction?
21	Were there moments during the virtual environment experience when you felt completely focused on the task or environment?
22	How easily did you adjust to the control devices used to interact with the virtual environment?
23	Was the information provided through different senses in the virtual environment (e.g., vision, hearing, touch) consistent?
24	Average

Table 4.11: Text of Presence questionnaire items for Tables 4.8 to 4.10 and Figures 4.16 to 4.17.

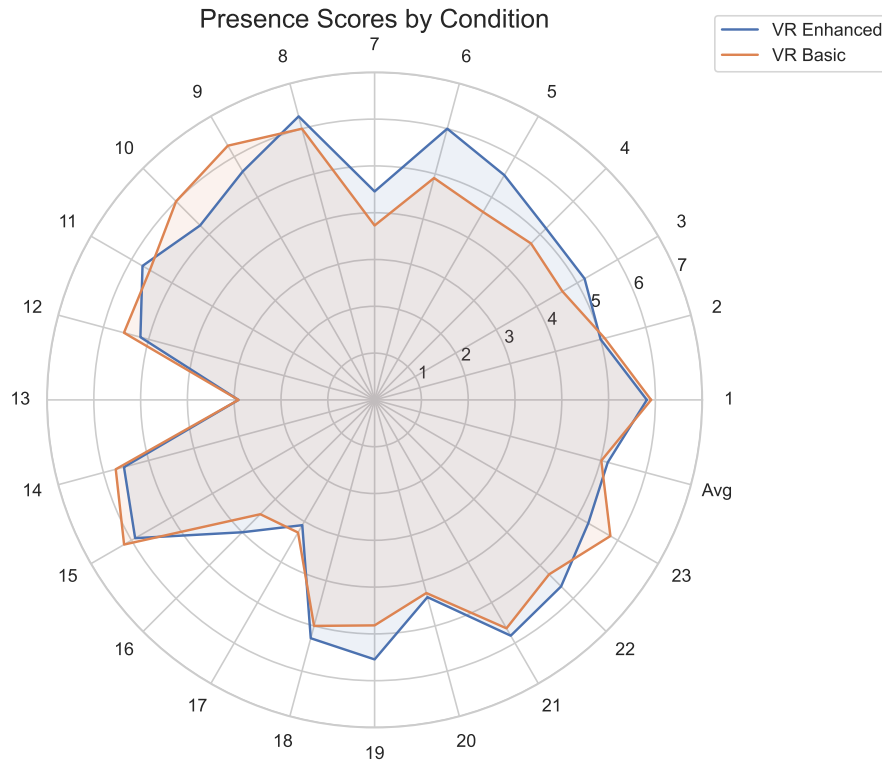


Figure 4.17: Presence radar plot.

- **Spatial Memory:** Working in VR, with respect to being able to recall information, I found being able to place objects anywhere in the environment:
- **Spatial Semantics:** Working in VR, with respect to being able to recall information, I found viewing individual function call graphs and control flow graphs:
- **Signalling:** Working in VR, with respect to formulating a solution to the problem, I found the visual callouts of function capabilities:
- **RE Tools:** Working in VR, with respect to formulating a solution to the problem, I found the text-based information (such as disassembly, string listings, etc.):
- **User Org[anization]:** Working in VR, with respect to formulating a solution to the problem, I found being able to place objects anywhere in the environment:

We provide a statistical overview of the data in Table 4.12 and visualization in Figure 4.18. Spatial Memory has the highest average rating and near-lowest standard deviation, indicating a slight preference for that affordance over others.

Condition	Statistic	Incr. Formalism	Spatial Memory	Spatial Semantics	Signalling	RE Tools	User Org.
VR Enhanced	mean	3.0	4.36	2.91	3.91	3.73	4.0
	std	1.34	1.03	1.51	0.94	1.19	1.26
	median	3.0	5.0	3.0	4.0	4.0	4.0
	min	1.0	2.0	1.0	2.0	2.0	1.0
	max	5.0	5.0	5.0	5.0	5.0	5.0
	count	11.0	11.0	11.0	11.0	11.0	11.0

Table 4.12: Statistical overview of ratings of affordances in VR (on scale 1-5).

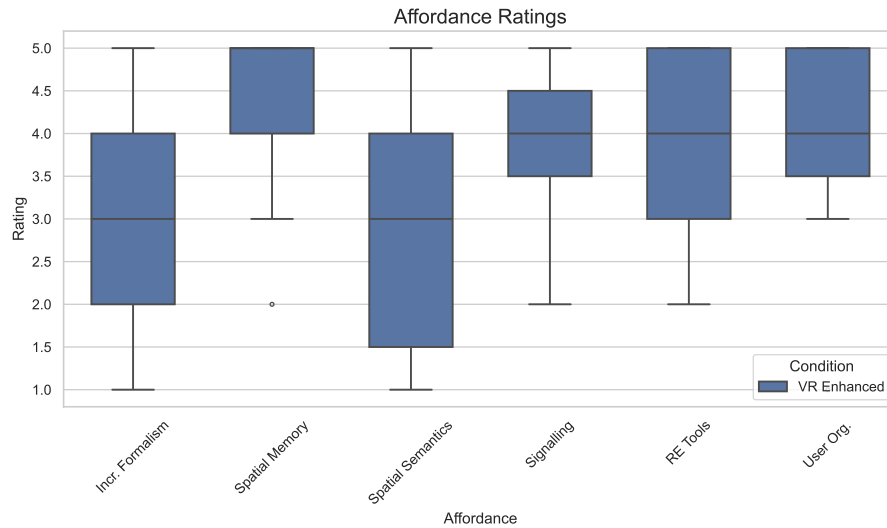


Figure 4.18: Affordance ratings box plot.

VR Log Measures

We used volume and area to measure how much of the VR space each participant used. Primarily, we examined the space used by the artifacts (slates and graphs) spawned by the participant. For VR Enhanced, we calculated volume using the center positions of each item by considering, conceptually, each position as a vertex of an arbitrary 3D object, and then using the Convex Hull algorithm (in Python, `scipy.spatial.ConvexHull`) to find the volume of that “object.” We calculated these volumes as used by the participant during answering each question and over answering all questions; we included all positions sampled (at the previously-mentioned rate of 2 samples per second) between the start and finish time of each question and then over all questions. This process gives us the fullest extent of space used by the artifacts in a transient environment. For VR Basic, we calculated area instead of volume, because all items were placed in the pre-determined positions as described in Section 4.1.2, where each item was placed at the same height (Y-axis). Thus we projected the coordinates to the X-Z plane and calculated the area used (for each question and over all questions) again using Convex Hull. Table 4.13 provides a statistical overview of the data, which is further visualized in Figures 4.19 and 4.20.

Further, we provide a visualization of the average number of items (slates and graphs) open at a time for each question and over all questions in Figure 4.21. Averaged across all participants, those in VR Enhanced consistently had more items open than VR Basic, but proportionally similar between the conditions.

4.2.2 Analysis

To evaluate differences across experimental conditions, we adopted a null hypothesis significance testing framework. For each comparison, the null hypothesis (H_0) posits that there is no difference between groups. We tested whether the observed differences could be attributed to random variation alone, using a significance threshold of $\alpha = 0.05$. Where results were not

Condition	Statistic	Over All Questions	Q1	Q2	Q3	Q4	Q5
VR Enhanced	mean	48.06	12.66	3.35	27.37	10.97	6.17
	std	93.8	37.86	5.13	81.87	21.38	13.24
	median	7.7	0.11	1.19	1.46	4.19	1.76
	min	3.72	0.01	0.06	0.27	0.31	0.0
	max	300.71	126.42	17.09	274.11	74.71	44.88
	count	11.0	11.0	11.0	11.0	11.0	11.0
VR Basic	mean	10.98	10.71	7.67	7.78	10.43	8.33
	std	2.78	0.67	2.71	2.28	2.86	3.14
	median	11.38	10.98	6.75	6.75	11.38	6.75
	min	4.65	9.71	4.65	4.65	4.65	4.65
	max	13.67	11.17	11.38	11.19	13.67	12.28
	count	11.0	4.0	11.0	11.0	10.0	11.0

Table 4.13: Statistical overview of volumes (VR Enhanced) and areas (VR Basic) occupied by graphs and slates.

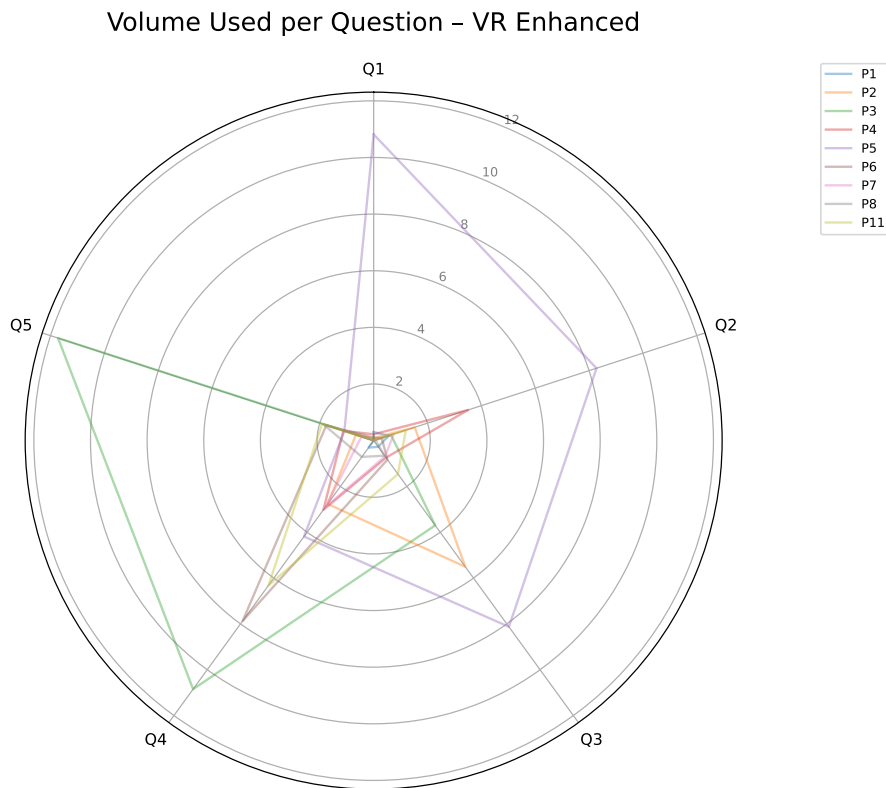


Figure 4.19: Volume used per question for each participant in the VR Enhanced condition; two participants were excluded due to being outliers.

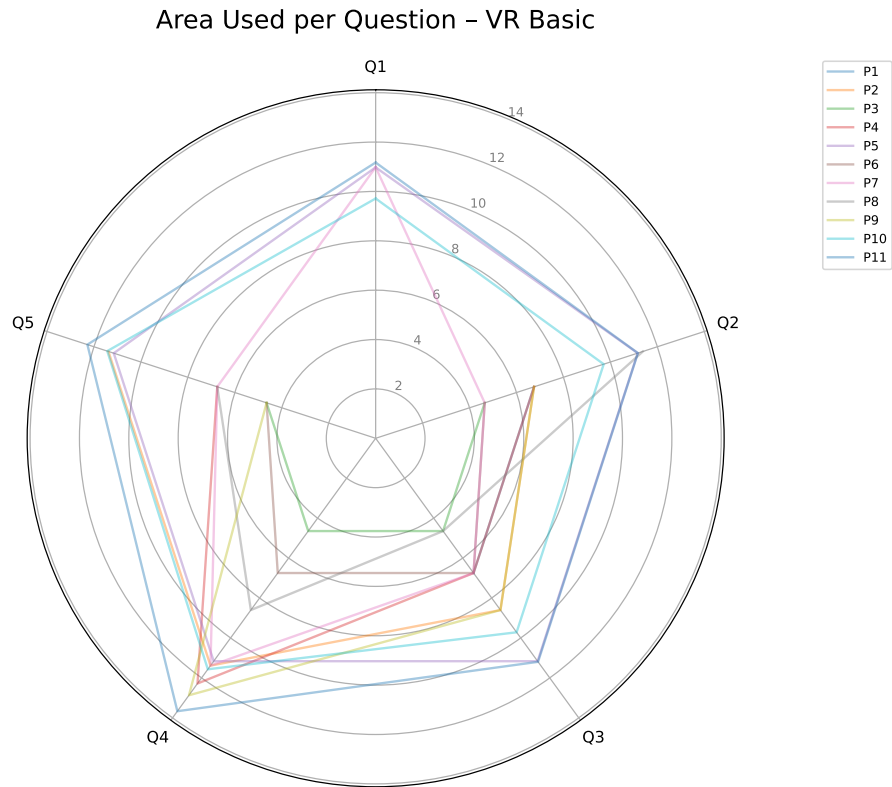


Figure 4.20: Area used per question for each participant in the VR Basic condition.

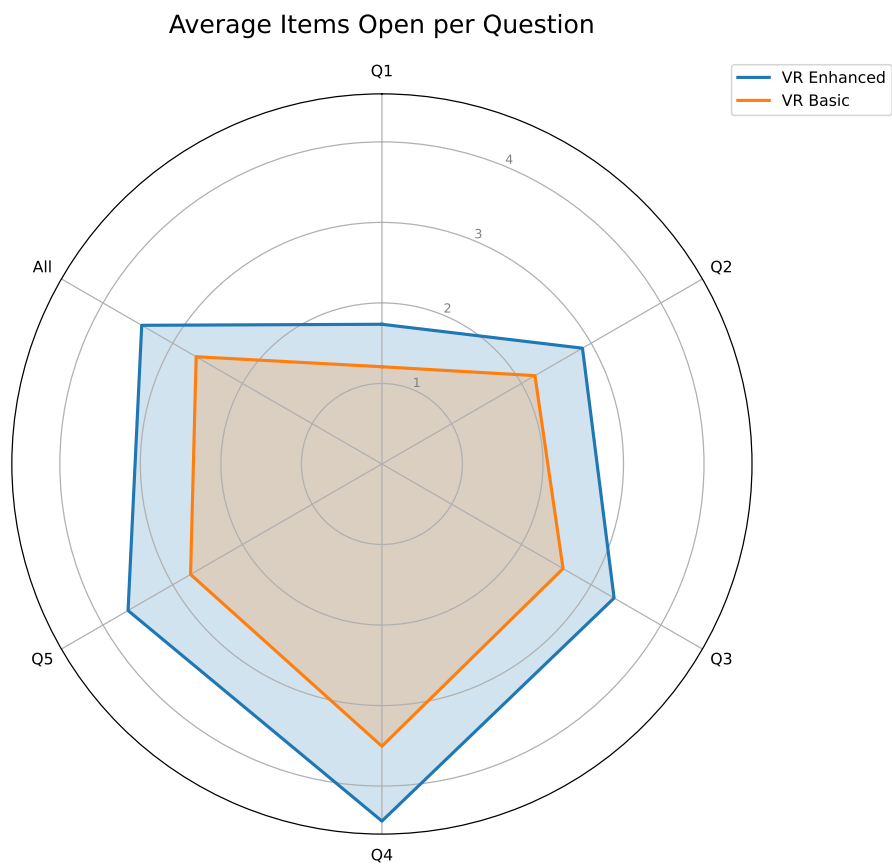


Figure 4.21: Average number of items open per question.

statistically significant, we report that we failed to reject the null hypothesis. This does not imply equivalence, only that our data do not provide sufficient evidence to conclude a difference exists.

To choose our tests, we checked our data for normality and equal variances. In Table 4.1, we listed a total of 37 variables that were collected in each of the three conditions (pre-survey, accuracy, time, TLX, SUS) plus 24 collected in the two VR conditions (Presence), for a total of 159 variables. We performed a Shapiro-Wilk test for normality ($\alpha = 0.05$) on those and found that 82 (52%) of them were not normally distributed. For the 61 variables collected in at least two conditions each (pre-survey, accuracy, time, TLX, SUS, Presence), we performed Levene's test on those and found that all have equal variances. Under these circumstances, in the remainder of this section, we predominantly use non-parametric tests for differences: Kruskal–Wallis H test for comparisons involving three groups and Mann–Whitney U test for pairwise comparison. In cases where the data is both normally distributed and have equal variances, we used parametric tests: One-way ANOVA (three groups) and Independent t-test (pairwise). For correlations, we used Pearson (parametric) or Spearman's rank (non-parametric) depending on the data.

Participant Analysis

To assess whether prior experience varied significantly across the experimental conditions, we conducted Kruskal-Wallis testing on participants' self-reported expertise in six relevant domains (as summarized in Table 4.2). We chose this test after we found that half of the pre-survey data variables were not normally distributed but all did have equal variances, per the Shapiro-Wilk test and Levene's test, respectively. No statistically significant differences were found in any category (all $p > 0.15$; see Table 4.14). We therefore fail to reject the null hypothesis of no difference. Three categories had higher variation (VR usage, VR development, and computer architecture) between groups while the remaining three categories had very low variation, which suggests differences in the groups, but they are not statistically significant. These results suggest that participants were comparably experienced across all three conditions, supporting the internal validity of between-group comparisons.

Test	SW Dev	Comp Arch	Compiler Const	Binary RE	VR User	VR Dev
K-W H(2,30)	0.42	3.09	0.13	0.42	3.13	1.89
K-W p-value	0.81	0.21	0.94	0.81	0.21	0.39

Table 4.14: Kruskal-Wallis H-stat and p-value for the three conditions for each pre-survey experience rating.

Hypothesis 1: Participants will have better task completion metrics with certain levels of affordances in VR.

The three conditions provided levels of affordances in VR: none (Web); minimal base-line affordances (VR Basic); and additional affordances (VR Enhanced). Participants' task performance was measured by correctness (percent of correct answers to the five scaffolded questions) and total task duration. Across conditions, performance distributions were relatively flat with no evidence of floor or ceiling effects. The individual question correctness variables were not normally distributed (per the Shapiro-Wilk test) but the average correctness was; and the majority of question duration variables were normally distributed, but not all. We used non-parametric tests for our analysis below (we also ran parametric tests, which showed the same significance results).

Regarding accuracy: We observed no statistically significant difference in overall accuracy between any of the conditions. Mean accuracy in the Web condition was 67.3% (SD = 26.7); 65.5% (SD = 27.0) in VR Basic; and 67.3% (SD = 28.7) in VR Enhanced. Among the three conditions, Kruskal-Wallis showed $H = 0.06$, $p = 0.97$ for average questions correct (for individual questions, p values ranged from 0.38 to 0.89). We therefore fail to reject the null hypothesis of no difference.

Regarding duration: Participants in the Web condition completed the task significantly faster than participants in either VR condition. Mean duration in the Web condition was 221.1 seconds (SD = 52.6); 267.2 (SD = 76.8) in VR Basic; and 273.6 (SD = 88.6) in VR Enhanced. Among the three conditions, Kruskal-Wallis showed $H = 2.76$, $p = 0.25$, hinting at a difference. Pairwise Mann-Whitney tests revealed no meaningful difference between the two VR conditions ($U = 63.0$, $p = 0.90$), but between VR Enhanced and Web, $U = 85.0$ and $p = 0.11$;

and between VR Basic and Web, $U = 30.0$ and $p = 0.05$. In this case, we do reject the null hypothesis of no difference (pairwise) between each VR and Web condition.

Although VR users did not achieve higher correctness, they engaged in longer sessions, possibly due to richer exploration or the impact of novelty. No performance advantage was observed for enhanced affordances. Compared to the Web condition, VR did not impose a penalty in correctness and only a small penalty in duration (22.3% higher average).

While the quantitative results do not show a substantial performance difference across conditions, participant quotes suggest that Enhanced VR created a more engaging and confidence-building environment, which may not have translated into measurable gains but was meaningful to participants. In VR Enhanced, participants reflected, “This was fun. I would rather use this than a desktop app,” and “I like the flowers in the environment. Fun touch.” Others in VR Basic said, “I’m a kinesthetic learner and I really appreciated being able to move in this environment,” and “[I] liked viewing functions in VR; [it’s] better than Ghidra.” These quotes suggest that embodied interaction and spatial affordances supported a sense of progress or agency, even if that didn’t yield significantly higher accuracy. Conversely, some Web participants expressed a desire to externalize in physical space, e.g., “I would prefer to draw on paper.” These quotes imply that participants perceived performance benefits in VR conditions, particularly Enhanced, but these perceptions did not translate into measurable performance improvements, perhaps due to other confounding factors (e.g., VR learning curve).

Hypothesis 2: Cognitive load will vary by level of affordances in VR.

Once again, the three conditions provided levels of affordances in VR: none (Web); minimal baseline affordances (VR Basic); and additional affordances (VR Enhanced). We tested NASA-TLX scores using Shapiro-Wilk, finding that seven of the 21 variables (across all conditions) were not normally distributed, so we used non-parametric tests (parametric tests also gave the same significance results).

For the TLX Average, Kruskal-Wallis showed $H = 4.379444$, $p = 0.111948$. Pairwise, there was no significant difference between VR Basic and Enhanced conditions; Mann-Whitney showed $U = 65.0$, $p = 0.7927$; however, between VR Enhanced and Web, $U = 37.0$ and $p =$

0.13; and between VR Basic and Web, $U = 30.0$, $p = 0.05$. The difference is attributable to a single question, “Performance: How successful were you in performing the task? How satisfied were you with your performance?”, where 0 is most successful (lowest load) and 100 is least successful (highest load). No other question variables, when tested with Kruskal-Wallis, had a p-value lower than 0.57, and excluding the “perceived success” question, there is no significant difference in any individual TLX items, or TLX Average, between conditions.

Focusing on the single “perceived success” variable, cognitive load was much worse in the Web condition (average 65.2) than VR Basic (average 38.5) or VR Enhanced (average 30.4). Pairwise, there was again no significant difference between VR Basic and Enhanced conditions; Mann-Whitney showed $U = 53.0$, $p = 0.64$; however, between VR Enhanced and Web, $U = 19.5$ and $p = 0.01$; and between VR Basic and Web, $U = 29.5$, $p = 0.04$. Taking the data at face value, there was significantly more cognitive load in the Web condition than either VR condition (we do reject the null hypothesis).

Participants in the Web condition often verbalized cognitive strain, stating “There’s a lot going on here,” and “I should have kept track of things better.” However, we acknowledge that it’s possible that participants simply misunderstood this question as it was posed, and that its value is an anomaly. Irrespective of condition, informally-observed cognitive load varied greatly amongst participants depending on how they responded to solving complex problems, performing under pressure, and information overload. Some participants heavily depended on the notepad, one uttering “I’m just going to write this down. I can’t keep track of all this stuff while doing all this,” and others took it as a challenge to avoid using it, one stating “[I’m] proud of myself; a lot of mental math I had to not mess up.”

Hypothesis 3: Participants will rate certain affordances in VR (e.g., user organization, spatial memory, signalling) higher than others in terms of usefulness.

Participants in the VR Enhanced condition rated six specific affordances using a 5-point Likert scale. Because we are comparing multiple variables measured on the same participants, we need a repeated measures test. The Shapiro-Wilk test revealed that four of the six

variables were normally distributed, so we used the Friedman test instead of repeated measures ANOVA. The Friedman test showed a significant omnibus difference across affordances ($\chi^2(5) = 12.39, p = 0.030$), but Bonferroni-corrected Wilcoxon tests did not yield significant pairwise differences. In this case, we weakly reject the null hypothesis: there is a significant difference across all affordances, but the analysis does not clearly reveal which affordance(s) would be most useful with statistical significance.

We asked about free object placement in two ratings, Spatial Memory (“Working in VR, with respect to being able to recall information, I found being able to place objects anywhere in the environment”) and User Organization (“Working in VR, with respect to formulating a solution to the problem, I found being able to place objects anywhere in the environment,”) and these affordances were rated highest (but not significantly higher than other affordances). Several participants made unsolicited positive (and constructive) comments about spatial manipulation and layout control, such as, “I enjoyed the spatial capabilities,” and “This is really cool and the way this moves is really seamless.” As a counterpoint, one participant remarked that they found it difficult to remember where the slates were due to their ability to move them freely. However, even without corrected significance in pairwise comparisons, this trend suggests the perceived importance of free object placement.

Hypothesis 4: Participants who have better performance in the immersive environments will have used more of the VR space.

We analyzed the volumes we calculated as described in the data overview for the VR Enhanced condition. Per Shapiro-Wilk testing, none of the volume variables were normally distributed, so we used Spearman’s Rank to test correlation. In VR Enhanced, volume used by participants’ artifacts in the environment is only weakly correlated, sometimes negatively, with correctness for each individual question ($\rho = -0.32$ to 0.15). Volume used over all questions is also only weakly correlated with average correctness ($\rho = 0.12$). This is essentially noise. Volume used was generally correlated with time taken with the range $\rho = 0.15$ to $\rho = 0.85$. This finding may be due to these participants using a deliberate process to thoroughly investigate all applicable information, opening up many slates in the process.

We further analyzed the areas calculated for VR Basic; none were normally distributed per Shapiro-Wilk testing, so we used Spearman's Rank to test correlation. Area used was negatively correlated with correctness for Q2 ($\rho = -0.48$), Q3 ($\rho = -0.51$), Q5 ($\rho = -0.48$), and overall ($\rho = -0.33$). Perhaps this correlation is due to some participants' uncertainty in where to find the information they needed, thus spawning more items than participants who better understood the task. Area used was somewhat correlated with time taken—positively and negatively—and without a discernible pattern, with the range $\rho = -0.47$ to $\rho = 0.57$.

We repeated these tests, for each question and for all questions, using average number of slates and graphs opened, maximum number of slates and graphs opened, and volume used by the HMD, and as above, we found no strong correlation (positive or negative) to question correctness or duration.

Other Interesting Findings

Although the primary hypotheses guided our main analysis, several additional findings provide further insight into key aspects of the system.

Revisiting the System Usability Scale results, while there were no significant differences between conditions, participants much more strongly agreed with positive-sentiment statements than negative-sentiment statements. Within each condition (comparing two measurements from the same participants), we performed Wilcoxon signed-rank tests to compare the two sets of non-normally-distributed values (the means of the positive ratings against the means of the negative ratings) to verify a statistically significant difference. We found a very strong difference; for VR Enhanced: $W = 0.00$, $p = 0.01$; VR Basic: $W = 1.00$, $p = 0.01$; and Web: $W = 0.00$, $p = 0.01$. These results indicate participants overwhelmingly found each condition highly usable, validating our user-centered design process and bolstering the internal validity of the study.

Most presence measures did not differ significantly between VR conditions. However, the VR Enhanced condition had notably stronger score on one measure, "How compelling was your sense of objects moving through space?" (mean = 6.00) compared to VR Basic (mean = 4.91),

with a Mann-Whitney U test yielding $p = 0.07$. This difference falls just short of being conventionally significant, but it aligns with qualitative observations that the additional affordances of the VR Enhanced condition contributed to an increased sense of embodied interaction.

While Pre-survey self-ratings were not significantly different between conditions, we checked for correlation between those ratings and correctness on the RE challenge questions. Using Pearson correlation (noting the data are not normally distributed), we found several moderately-high or high ($r > 0.60$) correlations:

- VR Enhanced: Software Development vs Q4, $r = 0.60$
- VR Basic: none
- Web: Software Development vs Q2, $r = 0.63$
- Web: Compiler Construction vs Q5, $r = 0.74$
- Web: Reverse Engineering vs Q5, $r = 0.85$
- Web: VR Developmet vs Q5, $r = 0.77$

These correlations suggest that prior domain expertise may be leveraged differently depending on the condition: in the more traditional Web condition, participants may rely more heavily on pre-existing schema and training to complete complex tasks.

In the VR conditions, high variation in volume or area used by participants suggests that participants developed unique workflows and that the system supports user-defined organizational strategies consistent with theories of spatial memory and external cognition.

Additional participant comments that stood out:

- “Controllers broke the immersion. Need total immersion with hand gestures and natural typing, or use augmented reality to augment a traditional desktop keyboard interface.” (VR Enhanced) — good design feedback for future improvements.
- “Worried to move [within physical space]. Don’t want to hit something.” (VR Enhanced) — exemplifies trepidations of novice VR users.

- “I was nervous that I wouldn’t understand the code but it was easy.” (VR Basic) — evidence that participants found the material approachable, which supports the study’s generalizability.

Discussion

While we found no strongly statistically significant performance gains between conditions, qualitative data and participant observations indicate that the spatial features in immersive VR influenced how participants interacted with the task. Participants explicitly referenced features as helpful, such as the ability to arrange artifacts in space, maintain visual context, and group related elements. These behaviors that reflect the construction of an “analytical workspace,” supporting the idea that immersion and spatial freedom change not only interaction but also the cognitive framing of the task.

The lack of performance differences may result from several interacting factors beyond statistical power alone. First, the task duration was relatively short in order to meet design goals stated in Section 4.1.4, which may have limited the time available for the benefits of spatial organization to manifest in measurable performance improvements. Second, the reverse engineering questions were designed to be moderately difficult and appropriate for a mostly novice participant pool, which may have limited the range of possible outcomes and suppressed between-group variability. Third, the high within-group variability in correctness, particularly in the VR conditions, suggests that individual differences in spatial ability, RE strategy, and prior technical experience likely influenced outcomes as much as or more than the interface condition itself.

The data hints at the existence of small effect sizes, however, our sample size ($n = 11$ per condition) is not sufficient to detect small (Cohen’s $d = 0.2$) effects. If the effect sizes are small, adding 7 participants per group (to meet our original goal of 18 per group) would still yield limited statistical power and would be unlikely to change the findings of the analysis. Conducting another power analysis reveals we would need 81 participants per condition to reliably detect small (Cohen’s $d = 0.2$) effects, and that is not practical under our current experimental design.

While our sample size limits the ability to detect small effects, the current sample size is sufficient for detecting medium and large effects and for evaluating the feasibility and qualitative impact of the immersive system.

4.3 Conclusion and Future Work

In this study, we sought to answer RQ6, *How do binary RE practitioners employ the affordances of embodied immersion in this system?* and RQ7, *Does this system improve outcomes in static binary RE tasks?* We carefully designed and executed a user study to test fundamental concepts found in our foundational research while working within practical limitations of our available pool of participants. We analyzed data from our logs and observations to answer RQ6 and performed standard statistical tests on collected data to answer RQ7.

While many hypotheses were not statistically supported, the process of designing, executing, and analyzing this study yielded insights that may benefit future research in immersive analytics, reverse engineering tools, and user-centered VR evaluation:

- Immersion alone does not guarantee performance gains. Although immersive interfaces offer rich interaction, immersion does not always improve accuracy or speed. Participants in the Web condition completed tasks more quickly, while VR participants showed greater engagement and exploratory behavior. Researchers should anticipate that immersive benefits may lie in different cognitive dimensions, such as reflection or memory support, not immediate performance gains.
- User-driven spatial layout was highly valued. Participants consistently praised free placement of objects as the most useful affordance. This finding reinforces the importance of supporting user agency in spatial organization. Rather than enforcing fixed metaphors, systems should allow users to create their own spatial logic.
- Standard surveys may miss subtle cognitive benefits: The NASA TLX did not differentiate VR conditions in workload, despite participant comments suggesting reduced mental juggling. As a post-activity survey, it remains an accessible but imprecise instrument

relying on the participant's reflection of the task, with more recent research turning to biometric measurements to assess cognitive load in real time when possible. Additionally, participants may have interpreted some TLX questions (e.g., perceived success) inconsistently. In our case, behavioral logging and qualitative data provided valuable insight beyond TLX.

- Individual strategy variation was substantial: Some participants heavily used spatial affordances while others ignored them (especially the graphs and sometimes the notepad). Performance did not neatly correlate with spatial usage, but strategy did. Studies should account for wide individual variation in how people use the tools provided, especially when spatial interaction is possible.
- Balancing realism and control is challenging but necessary: The constrained, scaffolded RE task was a practical solution to the constraints of this study, and it allowed comparison but limited ecological validity. Future studies should include more open-ended tasks that resemble real-world reverse engineering while retaining enough structure for meaningful evaluation, and use a within-subjects design if possible.
- Null results are still insightful: This study contributes constraints on the claims that can be made about immersive environments for analytical tasks. It shows where performance differences do not arise, but also documents what users appreciated and how they adapted. These findings help sharpen theoretical models and inform the design of follow-on experiments.

Taken together, these lessons reinforce the idea that user studies in immersive environments must blend controlled evaluation with openness to exploratory, emergent behavior. Tools that allow users to shape their spatial and cognitive environment, rather than dictate it, may offer the most value in complex analytic domains.

Chapter 5

An Experiment with AI

While CogBRE enables spatial organization and persistent externalization of information, the placement and structuring of visual artifacts remains manually controlled by the user. In this chapter, we explore the use of large language models (LLM) as visualization agents capable of dynamically generating and organizing immersive 3D representations of program artifacts, guided by patterns learned during training that reflect human perception and cognitive design principles. Our hypothesis is that LLMs contain sufficient latent, embedded cognitive and perceptual knowledge to construct visualizations that support sensemaking even without them being explicitly trained for this task. This chapter answers RQ8, *What are ways an AI-based agent could improve the user experience?*

This work is best understood as an early-stage design study with elements of formative evaluation. The contribution of this work includes the design and implementation of an LLM-integrated immersive RE system, the hypothesis that LLMs can act as cognitive-aware visualization agents without explicit training, and a pilot evaluation testing the cognitive plausibility of LLM-generated 3D visualizations. We focus specifically on the generation of function call graphs, a choice influenced by feedback from users in our second development iteration who, when prompted about desired major improvements, requested a 3D alternative to the 2D hierarchical graphs.

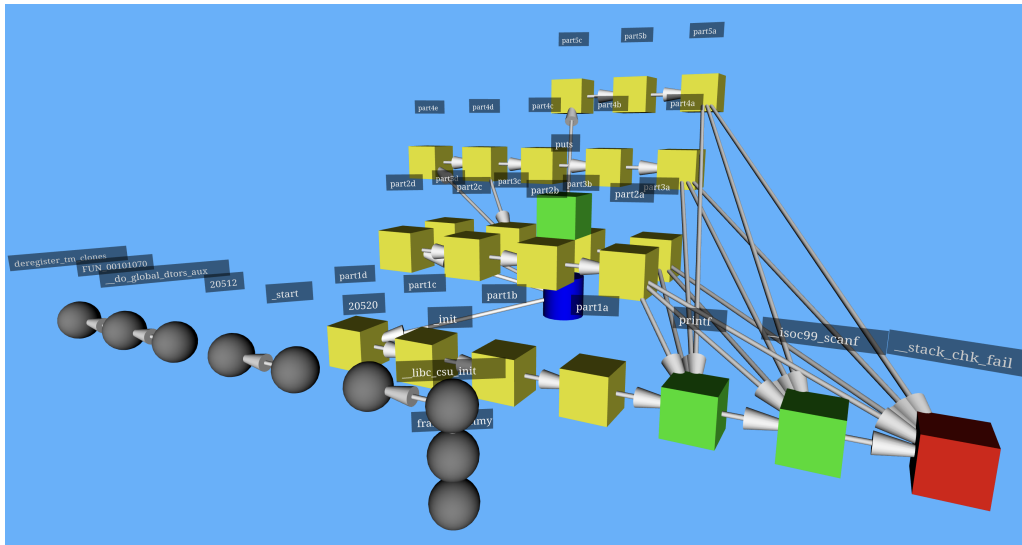


Figure 5.1: Call graph designed by an LLM after performing self-directed discovery of the v8 binary program employed in the user study discussed in Chapter 4. Note how it arranged the chains of functions (part1a..d, etc.) into rows and gave them a specific shape and color.

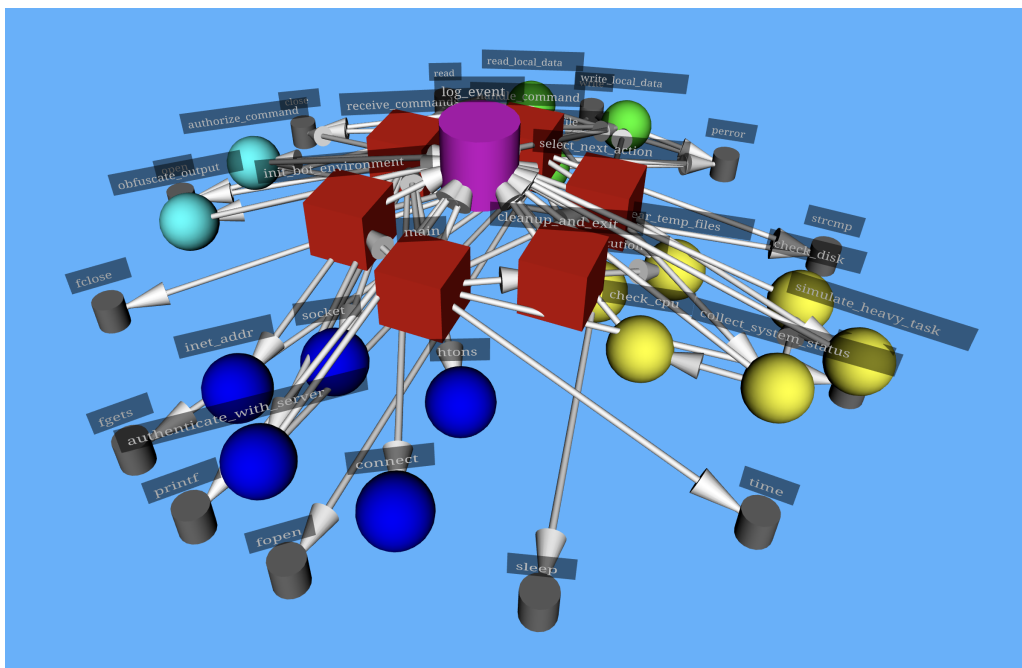


Figure 5.2: Call graph designed by an LLM after performing self-directed discovery of a binary program simulating a remote-controlled bot. Note overall layout in hemispherical shape with higher-level functions in the center radiating out to groups of related functions and system calls on the periphery.

5.1 Related Work and Motivation

Analytic tasks can benefit from effective visualization techniques that reduce cognitive load, enhance memory [108, 109], enabling efficient sensemaking. Our work explores such visualization for binary RE. It builds on foundational principles in visualization design and cognitive science to investigate whether LLMs can generate perceptually- and cognitively-aligned visualizations in an immersive RE environment.

Prior work in cognitive psychology and perceptual design provides principles for effective visualizations. Given the extensive inclusion of technical, scientific, and design documentation in typical LLM training corpora, it is reasonable to expect that patterns reflecting cognitive and perceptual design principles are embedded in LLM behavior. We do not claim that LLMs reason about perception or cognition in a principled manner; rather, we hypothesize that they can heuristically generate visualizations that reflect effective design patterns. Effective visualizations should align with user tasks [108] and should also exploit pre-attentive processing (color, shape, position), use spatial organization to support memory and reasoning, and apply grouping to reduce cognitive burden [109]. Visualizations should maximize information and minimize noise, avoid unnecessary complexity, and be quick to comprehend [110].

While current binary RE tools (e.g., Ghidra [7], IDA [6]) offer visualizations, they primarily provide only 2D graphs in conventional layouts that present a limited selection of qualities of the binary file per graph. Prior work in immersive analytics and software visualization has demonstrated the potential of spatial layouts, persistent visualizations, and embodied interaction to support sensemaking tasks [57, 66, 87, 89]. Recent advances demonstrate that LLMs can aid in source code tasks (e.g., Copilot, Codex), however, we have found no prior work that uses LLMs to autonomously create or organize 3D visual structures aligned with human cognitive affordances for the task of sensemaking in binary RE. LLMs may possess latent capabilities for generating useful, cognitively-aligned visualizations. Such visualizations could assist in reducing extraneous cognitive load, managing workspace organization, and supporting goal iteration within a 3D space that leverages embodied and external cognition.

5.2 System Overview: LLM-Augmented CogBRE

To test our hypothesis, we updated CogBRE by incorporating an intelligent agent driven by a large language model (LLM). At the core of this integration is the Model-Context-Protocol (MCP) framework [111], which structures interactions between the LLM, tools, and immersive environment. Figure 5.3 shows the updated architecture. This architecture enables the LLM to function as an informed collaborator, capable of tool use and immersive world manipulation.

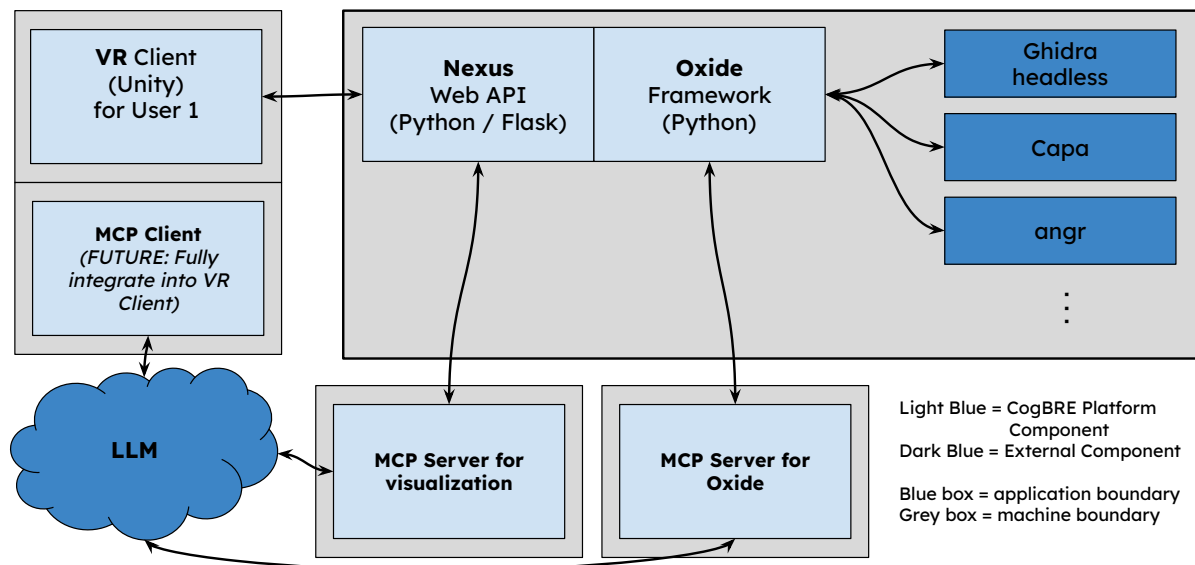


Figure 5.3: CogBRE System Architecture with LLM and MCP.

The system consists of six key components:

1. The CogBRE VR client, where the user interacts in immersive space with a tracked head-mounted display and hand controllers;
2. Nexus, a broker between the VR client and data providers;
3. Oxide, a modular RE analysis engine [97];
4. The MCP visualization server, which translates LLM directives into spatial commands for the VR client that are brokered by Nexus;
5. The MCP Oxide server, which provides tools to answer LLM questions about binary files through calls to the Oxide API (and indirectly, a suite of common RE tools); and

6. An MCP client, which mediates between the LLM and the rest of the system.

The latter three components are newly-added for the AI effort.

A sample scenario using this system plays out as follows: Upon system start, the MCP client provides the LLM the lists of visualization and RE tools. A user provides a prompt to the MCP client, which passes it on to the LLM: “Show me functions that write files.” The LLM determines what tools to use, likely requesting a list of capabilities of each function from Oxide, which would serve results from Capa [98]. The resulting function names are provided to the user in a list. The LLM may also query function relationships and contents in order to better summarize what each function does. Next, the LLM can build a visualization, such as a 3D call graph representation, highlighting the functions that write files, tagging them with summaries and showing their relationships to other functions, that is displayed in the VR environment for the user to further manipulate.

In the next section, we describe the design of our pilot evaluation.

5.3 Evaluation Design

In order to gain insight into our hypothesis that LLMs contain enough embedded knowledge to act as perceptual designers, we ran a pilot evaluation to rate visualizations generated by LLMs under varying conditions including level of guidance provided by the prompt. We employed subjective measures rated by humans (four graduate students studying binary reverse engineering) and objective measures calculated automatically.

We set limits on the pilot evaluation in order to reduce confounding variables, learn enough to inform future work, and respect the time of our raters. First, we limited the evaluation to LLM-generated function call graphs to enable meaningful measures (versus allowing the LLM to generate arbitrary geometric visualizations). Call graphs are a common, cognitively significant representation in RE; a call graph that reflects function roles, relationships, and flow aligns well with the cognitive affordances we explored in our prior research. Second, we varied three primary parameters, guidance level, model, and binary program, to create only 8 unique configuration prompts (more prompting details are in Section 5.4). Finally, we generated 5

visualizations per configuration (each set of 5 visualizations were generated with the same prompt) for a total set of 40 visualizations.

Each of the 4 raters received their own evaluation package with the LLM-generated visualizations (in X3D format) and the text of the LLM’s reasoning behind its design choices. Each package contained all 40 visualizations presented in a random order with the guidance condition and LLM hidden; raters knew the program file in order to compare it to ground truth about the program. Raters completed a spreadsheet of ratings in their own time using a desktop X3D viewer, rating each dimension (described next) for each visualization.

The subjective rating dimensions were inspired by Munzner, Ware, and Tufte as referenced in Section 5.1 plus a measure of correctness against ground truth. Raters provided a Likert scale rating of 1 (worst) to 5 (best) in each of these dimensions: *Clarity*: How easily can you understand what this visualization is showing? *Task Fit*: Would this help you answer questions about the program? *Spatial organization*: Are related elements grouped and arranged meaningfully in space? *Cognitive load*: What is the mental effort required to interpret this visualization? (1 = most effort; 5 = least) *Visual encodings*: Are size, shape, color, and position used effectively? *Correctness*: Does it lack obvious errors or misleading information?

Our objective metrics were chosen to quantify structural and perceptual properties of the LLM-generated call graphs. *Edge crossings*: sum of crossings detected when projecting the graph to the XY and XZ planes, which reflects layout clarity and directly impact the visual complexity of the graph. *Spatial dispersion*: the average pairwise distance between nodes, which serves as a proxy for grouping quality: tighter clusters may indicate more interpretable layouts. *Hierarchy depth*: reflects the structural complexity of the call graph and may influence how effectively the LLM captures the program’s control flow. *Color and shape diversity*: assess the variety of visual encodings applied to distinguish function roles or behaviors. *Average and standard deviation of edge lengths*: relate to the uniformity and spatial balance of the layout. While these metrics align with visualization design principles known to influence cognitive load, task efficiency, and visual clarity, we will use them primarily to measure *variation* between visualizations, not *quality* of the visualization.

5.4 Prompting Methodology

A key question in building an LLM-driven visualization system is whether to constrain the model to known visual conventions or to allow it to freely explore the space of possible visual encodings. Fine-tuning an LLM on human-designed visualization patterns may produce more cognitively-grounded outputs (layouts that align with spatial reasoning principles and visualization best practices). However, such fine-tuning risks narrowing the model’s generative capacity, potentially suppressing novel or serendipitous structures that may emerge from its broader training. Instead, we advocate for a middle ground: allow the LLM to remain flexible but guide it with prompt engineering. This design allows the model to explore novel spatial structures while remaining aligned with principles of human perception and task relevance.

We designed a prompt for each of the 8 configurations in two parts. Each starts with the language as shown in listing 5.1, giving consistent guidance to the LLM on how to query information for the call graph and directing which binary program to examine.

```
You are an assistant in an immersive virtual reality system designed to
help users reverse engineer binary programs.

The user operates in 3D space and can view visual artifacts such as call
graphs, control flow graphs, and text windows (slates) with function
listings.

Your overall goal is to use the available capabilities to help the user
understand how the binary works.

You should use the tools provided to understand what functions are in the
binary program, the capabilities of the functions, and review the
decompiled pseudo-source code to understand what key functions do. You may
need to make many tool calls.

The binary file has ID = <insert ID here>
```

```
After you have your best understanding of the program's structure and
purpose, build a 3D function call graph that is designed and organized in a
way to convey the most meaning to the user.
```

Listing 5.1: Common Prompt Introduction

The *guidance level* portion of each configuration has two possibilities, “low” and “high,” referring to the level of design guidance given in the LLM prompt to generate the call graph. The low guidance condition tests how the LLM behaves when only minimally framed with no cognitive design principles or layout heuristics: no mention of grouping, spatial layout, or design principles. This condition leverages only the latent design knowledge in the LLM. The remainder of the prompt for this condition is simply “Explain your reasoning.” The high guidance condition provides principles (not rules) that represent plausible heuristics the LLM may have seen during training. The remainder of the prompt for this condition is shown in listing 5.2.

```
To support the user's reasoning, try to:
- Group related elements spatially
- Use color or shape to distinguish different function types or behaviors
- Avoid unnecessary clutter or overlap
- Place important elements where they are easy to notice
- Label elements when that helps clarity
Explain your reasoning.
```

Listing 5.2: High Guidance Prompt Conclusion

Our high guidance condition is grounded in widely-accepted principles from visualization and cognitive science. We draw on work referenced in Section 5.1: Munzner’s nested design model [108] to align visualizations with user tasks; Ware’s perceptual theory [109], based on classical Gestalt principles, to account for preattentive processing, grouping, and cognitive load; and Tufte’s principles [110] to emphasize clarity, density, and visual economy. Together, these sources provide a well-rounded framework to assess how varying levels of design guidance affect the quality and cognitive plausibility of visualizations generated by an LLM in

immersive reverse engineering tasks. While this framework does not exhaust all possible dimensions of visualization design, it provides a practical and theoretically grounded basis for evaluating structural layout and cognitive alignment in the context of immersive static analysis.

Another configuration element is the binary program being examined. Raters examined two programs with differing characteristics to provide some generalization to our results. The first is “hexdump” from the Canonical Ubuntu [112] 20.04 x86-64 distribution, compiled with optimization and therefore obfuscated function names. The second is “v11,” a custom-built simulation of a potentially malicious bot written in C. This program attempts no obfuscation within the source and was compiled without optimization for x86-64, which carried over original function names to the binary file. The first program is a typical program as found “in the wild,” while the second was specifically designed to exercise our system and to facilitate understanding by the raters. Both programs are small, with tens of functions, as we found that larger programs would hit the LLMs’ context window and tokens-per-minute limitations. To establish ground truth, we provided raters with the source code and a traditional 2D call graph for each program.

The final configuration element is the LLM: OpenAI’s GPT-4.1 or o4-mini [113]. This variation gauges the impact of LLM selection on call graph creation. We chose a flagship chat model and a cost-optimized reasoning model to represent common types of commercial LLMs; we limited the selection to two models to keep this pilot evaluation manageable and cost-effective. The three parts in our configurations (guidance level, model, and program) each have two options, yielding the 8 distinct configurations we mentioned earlier.

5.5 Evaluation Results and Analysis

Using the previously-defined subjective and objective metrics, we evaluated the 8 configurations of LLM-generated call graphs (40 total). Table 5.2 shows the mean ratings and coefficients of variation (CV) for each of the 8 configurations. Subjectively, graphs for “v11” were rated higher than those for “hexdump,” likely because the “v11” graphs always had meaningful function names carried over from the unoptimized binary, while “hexdump” function

names were inconsistently inferred by the LLM; this finding may demonstrate the importance of semantic context in generating visualizations. Guidance level showed only a modest effect on average ratings (High = 3.00 vs. Low = 2.86), suggesting that LLMs benefit somewhat from informal perceptual design cues, but results were not conclusive. GPT-4.1 and o4-mini performed similarly on average.

Performing t-tests on the objective measures, we found that high guidance significantly increased shape diversity ($p = 0.01$), average edge length ($p = 0.02$), and edge length variability ($p = 0.05$). These metrics suggest more visually expressive and differentiated layouts under high guidance, though not always better-organized. GPT-4.1 visualizations had significantly more color diversity than those from o4-mini ($p < 0.01$), implying greater use of pre-attentive color channels, though this did not necessarily correspond to better ratings. All t-test results are shown in Table 5.1.

Rating Dim.	Low Guid.	High Guid.	p	GPT-4.1	o4-mini	p
Edge Crossings	31.30	69.05	0.28	33.75	66.60	0.35
Spatial Dispersion	5.22	6.76	0.09	6.05	5.94	0.90
Hierarchy Depth	3.65	3.05	0.24	3.10	3.60	0.33
Color Diversity	9.70	8.25	0.29	11.45	6.50	0.00
Shape Diversity	2.65	3.70	0.01	3.35	3.00	0.38
Average Edge Len	3.45	5.11	0.02	4.33	4.23	0.89
Std Dev Edge Len	1.56	2.51	0.05	2.12	1.95	0.72

Table 5.1: Objective measures across 40 graphs.

To assess variability, we first considered the objective measures and computed a composite diversity score using the average of normalized values as shown in Table 5.2. The most diverse configuration was “v11” + high guidance + o4-mini, while the least diverse was “hexdump” + low guidance + GPT-4.1. Notably, when considering subjective ratings, the same configuration produced both the highest- and lowest-rated graphs for both programs (“v11” + high guidance

+ GPT-4.1 and “hexdump” + high guidance + o4-mini). Those outputs, along with the basic 2D call graphs of each program, are shown in Figures 5.4 through 5.9.

Subjectively, raters preferred graphs that used spatial grouping, minimized clutter, and replaced or augmented obfuscated names with inferred semantic names. These observations suggest that LLMs include some latent alignment with perceptual and task-relevant design goals, even without explicit visualization training.

Program	Guidance Level	LLM	Avg Subj Rating Mean 1-5	Avg Subj Rating CV	Composite Obj Score Mean 0-1	Composite Obj Score CV
hexdump	high	gpt41	3.15	0.12	0.29	0.2
hexdump	high	o4mini	2.67	0.24	0.16	0.31
hexdump	low	gpt41	2.47	0.2	0.24	0.3
hexdump	low	o4mini	2.72	0.13	0.21	0.43
v11	high	gpt41	2.96	0.21	0.33	0.32
v11	high	o4mini	3.22	0.15	0.46	0.52
v11	low	gpt41	2.91	0.13	0.33	0.21
v11	low	o4mini	3.35	0.05	0.21	0.26
ALL hexdump			2.75	0.18	0.22	0.36
ALL v11			3.11	0.15	0.33	0.46
ALL High Guidance			3.0	0.18	0.31	0.54
ALL Low Guidance			2.86	0.16	0.25	0.34
ALL gpt41			2.87	0.18	0.3	0.27
ALL o4mini			2.99	0.17	0.26	0.66

Table 5.2: Mean and Coefficient of Variation for Subjective and Objective Ratings.

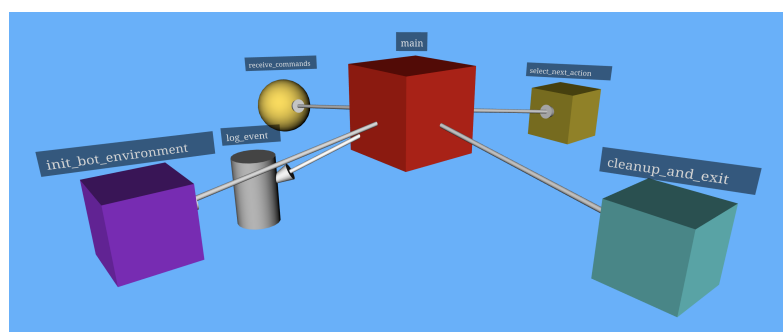
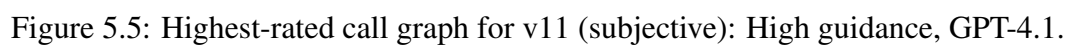
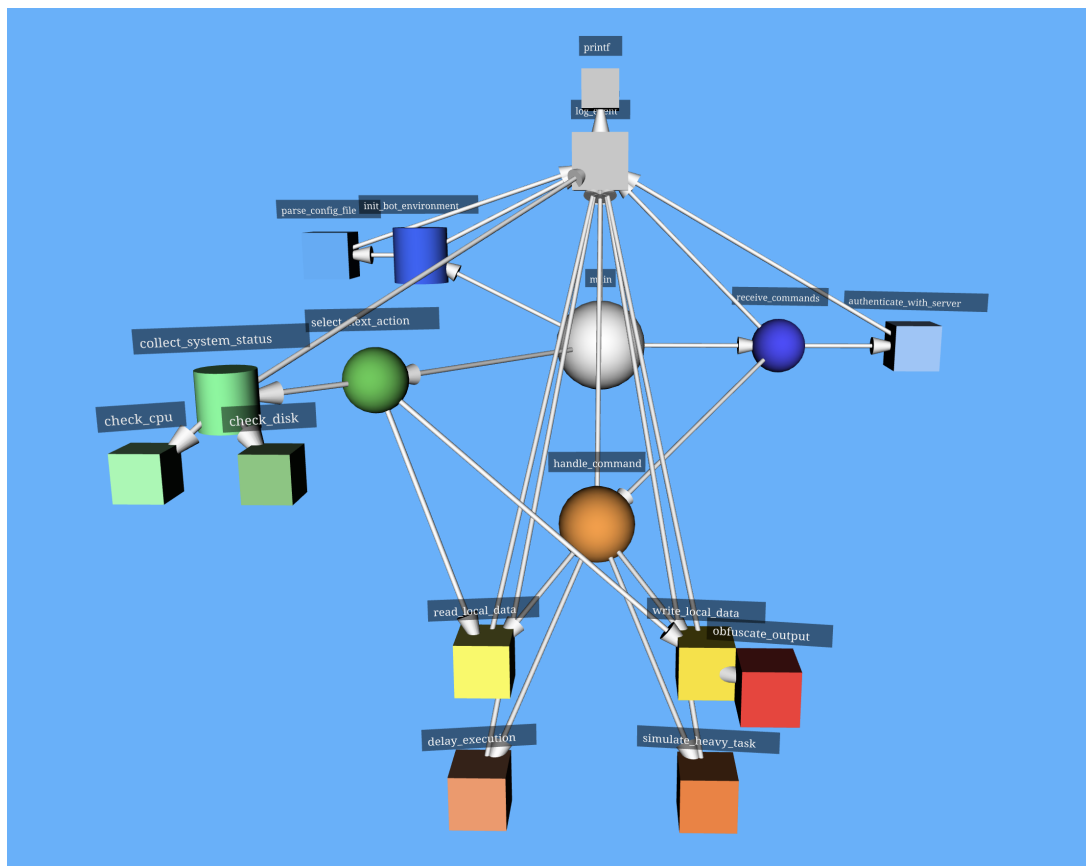
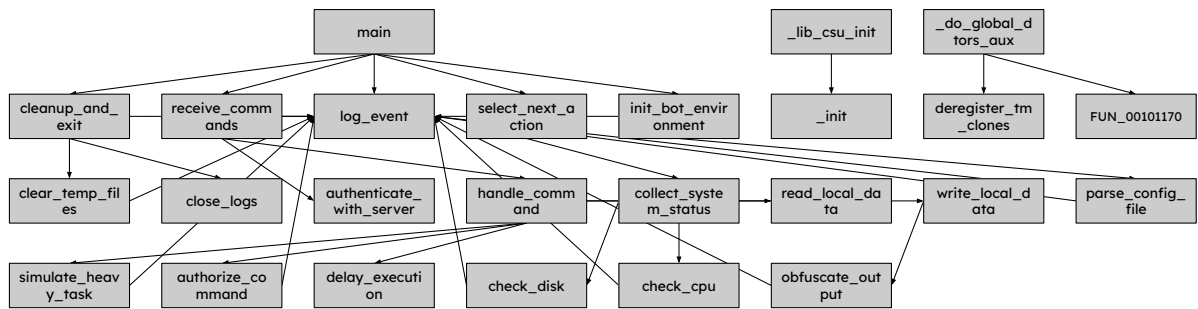


Figure 5.6: Lowest-rated call graph for v11 (subjective): High guidance, GPT-4.1.

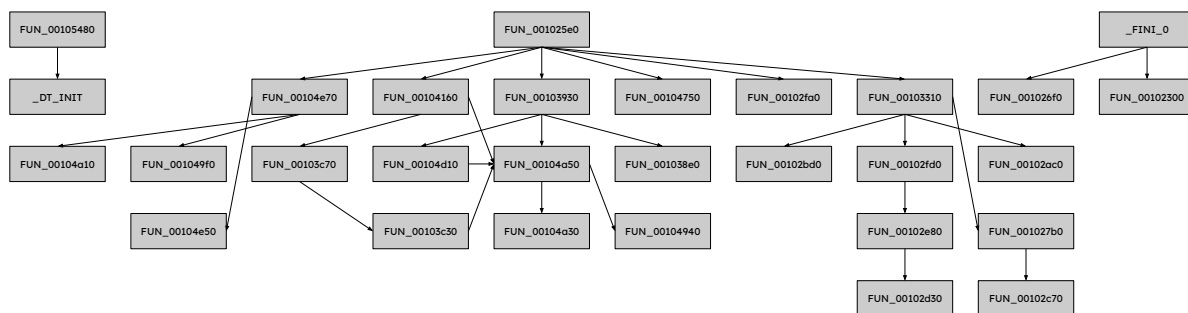


Figure 5.7: Basic 2D call graph for hexdump (without external calls).

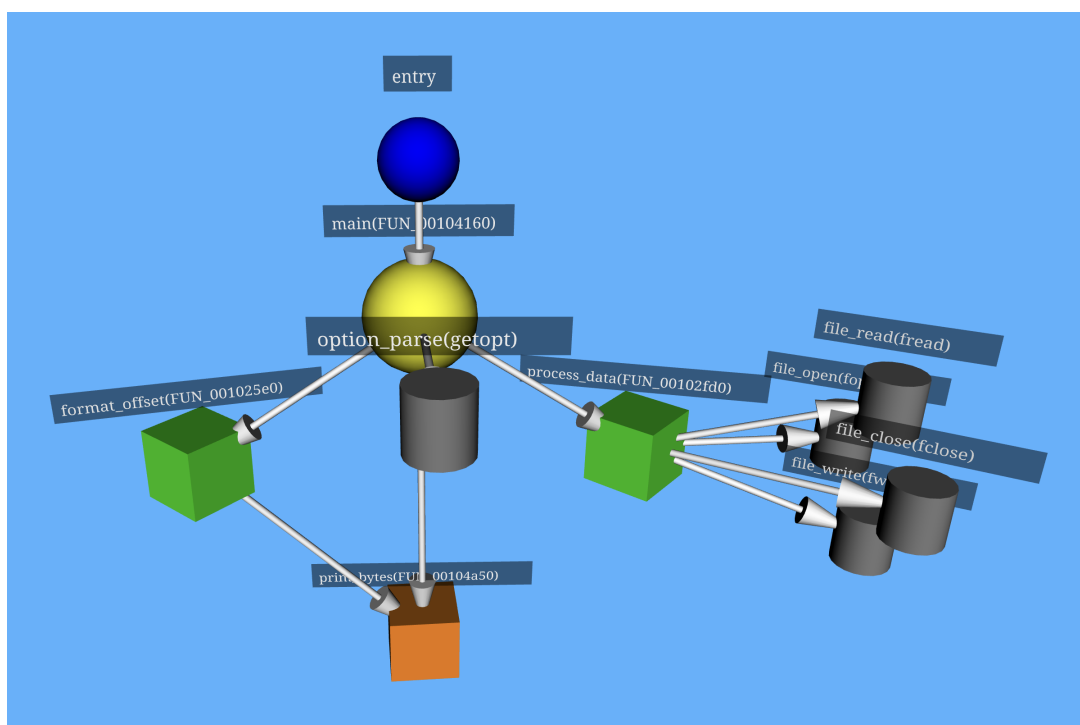


Figure 5.8: Highest-rated call graph for hexdump (subjective): High guidance, o4-mini.

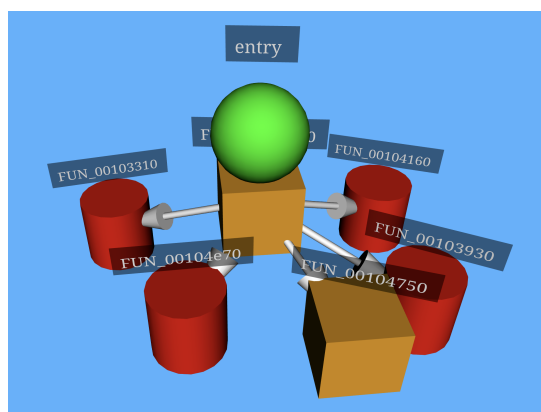


Figure 5.9: Lowest-rated call graph for hexdump (subjective): High guidance, o4-mini.

5.6 Conclusion

In this chapter, we answered RQ8, *What are ways an AI-based agent could improve the user experience?* We focused on generating 3D function call graphs as the basis of a formative pilot. While our pilot evaluation does not exhaust all possible dimensions of visualization design, it does support the hypothesis that LLMs can act as visualization agents to produce cognitively-influenced 3D layouts that reflect perceptually-grounded and task-aligned principles. Many generated call graphs exhibited meaningful spatial grouping, appropriate use of color and shape, and semantically-enriched labeling.

However, there are limitations that must be overcome before this approach is practical for routine use. There are currently severe limitations on the size of binary file supported by this approach; programs having more than tens of functions hit context window and token rate limitations quickly. Additionally, each call graph required 10 to 300 seconds to generate, which does not support true interactivity. The largest limitation is the variability in results when the LLM is allowed to determine what information to query about the program on each run; we could have provided the LLM a consistent set of information about the program on each run, but we wanted to let the LLM execute an integrated investigative process to discover and visualize.

Potential future work includes several avenues. First, integrate the MCP client into the VR interface for seamless prompting and iteration including speech-to-text and text-to-speech. Second, explore an actor-critic approach where a secondary model evaluates generated visualizations for cognitive fit. Third, determine an iterative method to support more complex binary programs. Finally, to address the variation in results, rather than add fixed limitations on the LLMs, use more implicit and explicit observations of the user’s cognitive and perceptual state to help focus the LLM’s generation of visualizations. Overall, our findings suggest that LLMs possess latent cognitive design capabilities but likely require a scaffolded approach to increase usability.

Chapter 6

Conclusion

Reverse engineering (RE) of binary software remains a complex, high-stakes task crucial to cybersecurity and software sustainment through tasks such as malware analysis and modernization efforts. Despite decades of tool development, the process remains cognitively demanding and human-centric, largely due to the compounded uncertainty introduced in the build process as source code becomes an executable program. Traditional tools provide critical information but fall short in directly supporting the analyst’s sensemaking process, particularly in navigating uncertainty, maintaining context, and forming and testing mental models of a program’s behavior.

This dissertation investigated whether embodied interaction through immersive virtual reality (VR) could offer a more cognitively-aligned interface for static binary RE. Drawing from several related disciplines, we hypothesized that affordances in VR such as spatial memory and semantics, external memory, and user organization, could reduce cognitive complexity and support more effective reverse engineering workflows. Our goal was not to replace existing tools, but to enhance the human-machine cognitive system through better alignment with human perception and cognition.

6.1 Summary of Contributions

This work makes several distinct contributions to the fields of reverse engineering, cognitive interaction design, and immersive analytics:

- **Theoretical Contribution:** Through a review of literature and synthesis across domains, we answered RQ1, identifying key characteristics of sensemaking and mental models in binary RE. We then addressed RQ2 by bridging these models to cognitive theories such as embodied cognition, external cognition, and cognitive load. In response to RQ3, we surveyed how immersive VR affordances have been employed in other analytic domains to enhance cognition. We synthesized these findings to inform our answer to RQ4: how immersive affordances might be leveraged to improve static binary RE.
- **System Design and Implementation:** In response to RQ5, we developed CogBRE, an immersive VR system for static binary reverse engineering, through a hybrid process combining Design Thinking and Cognitive Systems Engineering. The system integrates industry-standard RE tools into a spatial, embodied environment where users can spawn and manipulate visualizations of the information produced by multiple tools in a single expansive immersive environment.
- **Empirical Evaluation:** To address RQ6 and RQ7, we conducted a controlled user study comparing desktop and immersive VR conditions in performing a scaffolded binary RE task. While performance metrics showed minimal statistically significant differences between conditions, we observed lower cognitive load in immersive environments (particularly in mental demand and effort ratings), and users subjectively favored the immersive interface’s support for organization and spatial understanding.
- **Exploratory Use of AI Agents:** In response to RQ8, we conducted an exploratory study using LLMs to generate 3D call graphs of binary programs. Results showed promising potential for LLMs to act as cognitive partners, generating meaningful and novel layouts based on synthesized binary insights, which suggests a path toward semi-autonomous immersive RE agents.

- **Methodological Contribution:** This work offers a reproducible methodological framework for designing and evaluating cognitively aligned immersive tools. By explicitly grounding affordance design in cognitive theory and RE practice, and coupling qualitative feedback with quantitative metrics, we provide a template for similar efforts in other cognitively-intense technical domains.

6.2 Interpretation of Findings

Our evaluation yielded a mix of expected and surprising results. Objective task performance was not significantly better in immersive conditions, contrary to our primary hypothesis. However, this result aligns with findings in other immersive analytics research where spatial benefits are often qualitative, subjective, or long-term. Importantly, participants in immersive conditions reported lower mental workload and higher satisfaction. They consistently highlighted affordances like spatial arrangement, embodied navigation, and user-driven organization as valuable, even when they did not result in immediately faster or more accurate task outcomes.

These results support the notion that immersive systems, when designed with cognitive alignment in mind, can support sensemaking and reduce user burden. The immersive conditions did not merely add novelty; they provided new ways to manage complexity and maintain context, especially useful in the iterative, hypothesis-driven workflow of reverse engineering. Participants used the immersive space as a personalized analytical workspace, constructing spatial groupings and reference layouts that reflect an alternative cognitive framing of the RE task.

The LLM-based experiments further point to an emerging opportunity: intelligent agents that not only extract data, but also proactively generate visualizations aligned with human cognitive strategies. These experiments were preliminary, but demonstrate that AI and embodied interaction can co-evolve.

6.3 Limitations

This work has several important limitations. The user study sample size, while adequate for initial comparison, limits generalizability, particularly across experience levels. The reverse engineering tasks were carefully scoped to fit within a constrained set of evaluation parameters and may not reflect the full spectrum of real-world RE challenges.

Our VR system focused on spatial organization and cognitive affordances but did not support execution or dynamic analysis. As with many immersive systems, physical space constraints and hardware limitations may have influenced how fully participants utilized the environment. Furthermore, the integration of AI agents was limited to one-off generation of manipulatable artifacts but without further collaboration.

6.4 Future Work

Several promising directions emerge from this research:

- **Extended Studies and Longitudinal Use:** Future studies could involve professional reverse engineers performing deeper and more realistic tasks to assess whether immersive systems lead to improved understanding or reduced fatigue over time, even if initial task metrics remain flat.
- **Advanced Affordances and Mixed Modalities:** Incorporating eye tracking, gesture-based controls, voice input, haptic feedback, or measures such as electroencephalography (EEG) may reveal additional cognitive benefits. Mixed Reality (MR) or Augmented Reality (AR) systems could combine the flexibility of VR's extensive visualization capabilities with familiar desktop software and hardware, combining the best of real and virtual modalities.
- **Real-Time AI Collaboration:** Developing AI agents that collaborate in real-time, suggesting visualizations, workspace arrangements, or answering questions, could turn immersive RE into a conversational, co-creative process.

- **Broader Applications:** The methodology and findings may apply to other cognitively complex tasks beyond RE, such as cybersecurity forensics, algorithm behavior auditing, system-level debugging, or other domains where spatial reasoning and cognitive load are critical factors.
- **Integration with Educational Environments:** Immersive tools like CogBRE could be adapted for instructional use, offering students embodied experiences that align abstract concepts (e.g., control flow) with physical interaction, potentially improving retention and engagement.

6.5 Closing Remarks

This dissertation explored how making the abstract physical, through embodied interaction, spatial organization, and immersive visualization, can change how we reason about binary software. By aligning tool design with cognitive science and human-centered engineering, we demonstrated that it is more than what we see on the screen, but also how we inhabit and manipulate that information, that matters.

In an age of growing software complexity, AI assistance, and immersive interfaces, our findings suggest that the future of binary reverse engineering lies not only in smarter algorithms, but in smarter interfaces that empower the human mind to do what it does best: see patterns, build hypotheses, and make sense of uncertainty. This work is a step toward that future.

References

- [1] B. Shneiderman and R. Mayer, “Syntactic/semantic interactions in programmer behavior: A model and experimental results,” *International Journal of Computer & Information Sciences*, vol. 8, no. 3, pp. 219–238, Jun 1979. [Online]. Available: <https://doi.org/10.1007/BF00977789>
- [2] A. Bryant, R. Mills, G. Peterson, and M. Grimaila, “Software reverse engineering as a sensemaking task,” *Journal of Information Assurance and Security*, vol. 6, pp. 483–494, 01 2012.
- [3] P. P. Dudenhofer, “Modeling and automating the cyber reverse engineering cognitive process,” in *23rd Colloquium for Information Systems Security Education*, ser. CISSE ’19, 2019.
- [4] D. Votipka, S. Rabin, K. Micinski, J. S. Foster, and M. L. Mazurek, “An observational investigation of reverse Engineers’ processes,” in *29th USENIX Security Symposium (USENIX Security 20)*. USENIX Association, Aug. 2020, pp. 1875–1892. [Online]. Available: <https://www.usenix.org/conference/usenixsecurity20/presentation/votipka-observational>
- [5] X. Meng and B. P. Miller, “Binary code is not easy,” in *Proceedings of the 25th International Symposium on Software Testing and Analysis*, ser. ISSTA 2016. New York, NY, USA: Association for Computing Machinery, 2016, p. 24–35. [Online]. Available: <https://doi.org/10.1145/2931037.2931047>

- [6] Hex-Rays, “IDA Pro binary code analysis tool,” <https://hex-rays.com/ida-pro/>, 2023. [Online]. Available: <https://hex-rays.com/ida-pro/>
- [7] US National Security Agency, “Ghidra software reverse engineering suite of tools,” <https://ghidra-sre.org/>, 2023. [Online]. Available: <https://ghidra-sre.org/>
- [8] pancake, “Radare2: Libre Reversing Framework for Unix Geeks,” <https://www.radare.org/n/radare2.html>, 2023. [Online]. Available: <https://www.radare.org/n/radare2.html>
- [9] Y. David, U. Alon, and E. Yahav, “Neural reverse engineering of stripped binaries using augmented control flow graphs,” *Proceedings of the ACM on Programming Languages*, vol. 4, no. OOPSLA, pp. 1–28, nov 2020. [Online]. Available: <https://doi.org/10.1145/%2F3428293>
- [10] A. Maier, H. Gascon, C. Wressnegger, and K. Rieck, “Typeminer: Recovering types in binary programs using machine learning,” in *International Conference on Detection of intrusions and malware, and vulnerability assessment*, 2019.
- [11] E. Hollnagel and D. Woods, *Joint Cognitive Systems: Foundations of Cognitive Systems Engineering*. CRC Press, 2005. [Online]. Available: <https://books.google.vu/books?id=IwRHwOK2IzYC>
- [12] A. Mantovani, S. Aonzo, Y. Fratantonio, and D. Balzarotti, “RE-Mind: a First Look Inside the Mind of a Reverse Engineer,” in *31st USENIX Security Symposium (USENIX Security 22)*. USENIX, 2022.
- [13] P. P. Dudenhofer and A. R. Bryant, “Establishing a cognitive understanding of cyber reverse engineering tasks,” 2017.
- [14] S. Smits, “Enhancing binary analysis through cognitive load theory,” Master’s thesis, Arizona State University, 2022.
- [15] M. Wilson, “Six views of embodied cognition,” *Psychon Bull Rev*, vol. 9, no. 4, pp. 625–636, Dec. 2002.

- [16] T. Chandler, M. Cordeil, T. Czauderna, T. Dwyer, J. Glowacki, C. Goncu, M. Klapperstueck, K. Klein, K. Marriott, F. Schreiber, and E. Wilson, “Immersive analytics,” in *2015 Big Data Visual Analytics (BDVA)*, 2015, pp. 1–8.
- [17] T. Dwyer, K. Marriott, T. Isenberg, K. Klein, N. Riche, F. Schreiber, W. Stuerzlinger, and B. H. Thomas, *Immersive Analytics: An Introduction*. Cham: Springer International Publishing, 2018, pp. 1–23. [Online]. Available: <https://doi.org/10.1007/978-3-030-01388-2>
- [18] B. Ens, M. Cordeil, C. North, T. Dwyer, L. Besançon, A. Prouzeau, J. Liu, A. Cunningham, A. Drogemuller, K. A. Satriadi, and B. H. Thomas, “Immersive analytics 2.0: Spatial and embodied sensemaking,” in *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems*, ser. CHI EA ’22. New York, NY, USA: Association for Computing Machinery, 2022. [Online]. Available: <https://doi.org/10.1145/3491101.3503726>
- [19] H. G. Rice, “Classes of recursively enumerable sets and their decision problems,” *Journal of Symbolic Logic*, vol. 19, no. 2, pp. 121–122, 1954.
- [20] D. Crockford, *JavaScript: The Good Parts*. O’Reilly Media, Inc., 2008.
- [21] E. Mourão, J. F. Pimentel, L. Murta, M. Kalinowski, E. Mendes, and C. Wohlin, “On the performance of hybrid search strategies for systematic literature reviews in software engineering,” *Information and Software Technology*, vol. 123, p. 106294, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0950584920300446>
- [22] K. J. W. Craik, *The Nature of Explanation*. Cambridge: Cambridge University Press, 1943.
- [23] P. Johnson-Laird, *Mental Models*, ser. Cognitive science series. Cambridge University Press, 1983.
- [24] J. Preece, Y. Rogers, and H. Sharp, *Interaction Design: Beyond Human-Computer Interaction*, 5th ed. Hoboken, NJ: Wiley, 2019.

- [25] J. R. Anderson and C. Lebiere, *The atomic components of thought*. Mahwah, New Jersey: Lawrence Erlbaum associates, 1998.
- [26] G. A. Miller, “The magical number seven, plus or minus two: Some limits on our capacity for processing information.” *Psychological Review*, vol. 63, pp. 81–97, 1956. [Online]. Available: <https://doi.org/10.1037/h0043158>
- [27] R. Brooks, “Towards a theory of the comprehension of computer programs,” *International Journal of Man-Machine Studies*, vol. 18, no. 6, pp. 543–554, 1983. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0020737383800315>
- [28] K. A. Weigand and R. Hartung, “Abduction’s role in reverse engineering software,” in *2012 IEEE National Aerospace and Electronics Conference (NAECON)*, 2012, pp. 57–62.
- [29] I. Zayour and T. C. Lethbridge, “A cognitive and user centric based approach for reverse engineering tool design,” in *Proceedings of the 2000 Conference of the Centre for Advanced Studies on Collaborative Research*, ser. CASCON ’00. IBM Press, 2000, p. 16.
- [30] G. Klein, J. K. Phillips, E. L. Rall, and D. A. Peluso, “A data-frame theory of sense-making,” in *Expertise out of context: Proceedings of the Sixth International Conference on Naturalistic Decision Making*, R. R. Hoffman, Ed. Lawrence Erlbaum Associates Publishers, 2007, pp. 113–155.
- [31] F. E. Ritter, F. Tehranchi, and J. D. Oury, “ACT-R: A cognitive architecture for modeling cognition,” *WIREs Cognitive Science*, vol. 10, no. 3, p. e1488, 2019. [Online]. Available: <https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/wcs.1488>
- [32] M. K. Tennor, “Reverse engineering cognition,” MITRE Corporation, McLean, VA, Tech. Rep., 2015.

- [33] Z. D. Sisco, P. P. Dudenhofer, and A. R. Bryant, “Modeling information flow for an autonomous agent to support reverse engineering work,” *The Journal of Defense Modeling and Simulation*, vol. 14, no. 3, pp. 245–256, 2017.
- [34] M. Nyre-Yu, K. Butler, and C. Bolstad, “A task analysis of static binary reverse engineering for security,” in *55th Hawaii International Conference on System Sciences, HICSS 2022, Virtual Event / Maui, Hawaii, USA, January 4-7, 2022*. ScholarSpace, 2022, pp. 1–10. [Online]. Available: <http://hdl.handle.net/10125/79608>
- [35] I. Vessey, “Expertise in debugging computer programs: A process analysis,” *International Journal of Man-Machine Studies*, vol. 23, pp. 459–494, 1985.
- [36] M.-A. Storey, “Theories, methods and tools in program comprehension: past, present and future,” in *13th International Workshop on Program Comprehension (IWPC’05)*, 2005, pp. 181–191.
- [37] F. Détienne and F. Bott, *Software Design—Cognitive Aspects*. Berlin, Heidelberg: Springer-Verlag, 2001.
- [38] J. Siegmund, C. Kästner, S. Apel, C. Parnin, A. Bethmann, T. Leich, G. Saake, and A. Brechmann, “Understanding understanding source code with functional magnetic resonance imaging,” in *Proceedings of the 36th International Conference on Software Engineering*, ser. ICSE 2014. New York, NY, USA: Association for Computing Machinery, 2014, p. 378–389. [Online]. Available: <https://doi.org/10.1145/2568225.2568252>
- [39] J. Cowley, “Job analysis results for malicious-code reverse engineers: A case study,” 2014.
- [40] M. W. Eysenck and M. Brysbaert, *Fundamentals of Cognition 3rd Edition*. London: Routledge, 2018.
- [41] D. A. Norman, *Things That Make Us Smart: Defending Human Attributes in the Age of the Machine*. USA: Addison-Wesley Longman Publishing Co., Inc., 1993.

- [42] M. Scaife and Y. Rogers, “External cognition: how do graphical representations work?” *International Journal of Human-Computer Studies*, vol. 45, no. 2, pp. 185–213, 1996.
- [43] J. J. Gibson, “The theory of affordances,” in *Perceiving, acting, and knowing: Toward an ecological psychology*, R. Shaw and J. Bransford, Eds. Lawrence Erlbaum Associates, 1977, pp. 67–82.
- [44] D. Norman, *The Psychology of Everyday Things*. Basic Books, 1988.
- [45] T. van Gelder and R. F. Port, “It’s about time: an overview of the dynamical approach to cognition,” 1996.
- [46] A. M. Glenberg, J. K. Witt, and J. Metcalfe, “From the revolution to embodiment: 25 years of cognitive psychology,” *Perspectives on Psychological Science*, vol. 8, no. 5, pp. 573–585, 2013, pMID: 26173215. [Online]. Available: <https://doi.org/10.1177/1745691613498098>
- [47] L. Shapiro and S. Spaulding, “Embodied Cognition,” in *The Stanford Encyclopedia of Philosophy*, Winter 2021 ed., E. N. Zalta, Ed. Metaphysics Research Lab, Stanford University, 2021.
- [48] D. Kirsh, “Embodied cognition and the magical future of interaction design,” *ACM Transactions on Computer-Human Interaction*, vol. 20, no. 1, apr 2013. [Online]. Available: <https://doi.org/10.1145/2442106.2442109>
- [49] E. Hornecker, P. Marshall, and J. Jörn Hurtienne, “Locating theories of embodiment along three axes: 1st–3d person, body-context, practice-cognition,” in *CHI 2017 workshop on Soma-Based Design Theory*, 2017.
- [50] M. Ale, M. Sturdee, and E. Rubegni, “A systematic survey on embodied cognition: 11 years of research in child–computer interaction,” *International Journal of Child-Computer Interaction*, vol. 33, p. 100478, 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2212868922000174>

- [51] J. Sweller and P. Chandler, “Why some material is difficult to learn,” *Cognition and Instruction*, vol. 12, pp. 185–233, 1994.
- [52] J. Sweller, J. J. G. Van Merriënboer, and F. Paas, “Cognitive architecture and instructional design: 20 years later,” *Educational Psychology Review*, vol. 31, pp. 261–292, 06 2019.
- [53] F. Paas, A. Renkl, and J. Sweller, “Cognitive load theory: Instructional implications of the interaction between information structures and cognitive architecture,” *Instructional Science*, vol. 32, no. 1, pp. 1–8, Jan 2004. [Online]. Available: <https://doi.org/10.1023/B:TRUC.0000021806.17516.d0>
- [54] N. Hollender, C. Hofmann, M. Deneke, and B. Schmitz, “Integrating cognitive load theory and concepts of human–computer interaction,” *Computers in Human Behavior*, vol. 26, pp. 1278–1288, 11 2010.
- [55] D. Helgesson and P. Runeson, “Towards grounded theory perspectives of cognitive load in software engineering,” in *Psychology of Programming Interest Group (PPIG)*. Psychology of Programming Interest Group, Jun. 2021, psychology of Programming Interest Group Annual Workshop 2021 ; Conference date: 21-06-2021 Through 25-06-2021.
- [56] K. Brodman, *Brodmann’s Localisation in the Cerebral Cortex*. New York: Springer, 2006.
- [57] L. Lisle, K. Davidson, E. J. Gitre, C. North, and D. A. Bowman, “Sensemaking strategies with immersive space to think,” in *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, 2021, pp. 529–537.
- [58] C. Andrews, A. Endert, and C. North, “Space to think: Large high-resolution displays for sensemaking,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI ’10. New York, NY, USA: Association for Computing Machinery, 2010, p. 55–64. [Online]. Available: <https://doi.org/10.1145/1753326.1753336>

- [59] L. Lisle, X. Chen, J. E. Gitre, C. North, and D. A. Bowman, “Evaluating the benefits of the immersive space to think,” in *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, 2020, pp. 331–337.
- [60] M. Derksen, T. Kuhlen, M. Botsch, and T. Weissker, “Minimalism or creative chaos? on the arrangement and analysis of numerous scatterplots in immersive 3d knowledge spaces,” *IEEE transactions on visualization and computer graphics*, vol. PP, 03 2025.
- [61] J. Moloney, B. Spehar, A. Globa, and R. Wang, “The affordance of virtual reality to enable the sensory representation of multi-dimensional data for immersive analytics: from experience to insight,” *Journal of Big Data*, vol. 5, no. 1, 2018.
- [62] W. Cui, “Visual analytics: A comprehensive overview,” *IEEE Access*, vol. 7, pp. 81 555–81 573, 2019.
- [63] D. Gračanin, “Immersion versus embodiment: Embodied cognition for immersive analytics in mixed reality environments,” in *Augmented Cognition: Intelligent Technologies*, D. D. Schmorow and C. M. Fidopiastis, Eds. Cham: Springer International Publishing, 2018, pp. 355–368.
- [64] M. Billinghamurst, B. Thomas, J. Chahl, I. Bornkessel-Schlesewsky, M. Immink, and M. Schlewsky. (2019, Dec.) Enhancing human performance using virtual reality, wearable computing, cognitive neuroscience and mental training. [Online]. Available: <https://www.dst.defence.gov.au/sites/default/files/events/documents/9.0\%20Improving-cognitive-performance-UniSA-Billinghurst-Chaal.pdf>
- [65] M. Ahmadi, H. Bai, A. Chatburn, M. A. Najatabadi, B. C. Wünsche, and M. Billinghamurst, “Comparison of physiological cues for cognitive load measures in vr,” in *2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, March 2023, pp. 837–838.
- [66] A. Batch, A. Cunningham, M. Cordeil, N. Elmqvist, T. Dwyer, B. H. Thomas, and K. Marriott, “There is no spoon: Evaluating performance, space use, and presence with

- expert domain users in immersive analytics,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 26, no. 1, pp. 536–546, 2020.
- [67] M. Cordeil, A. Cunningham, T. Dwyer, B. H. Thomas, and K. Marriott, “Imaxes: Immersive axes as embodied affordances for interactive multivariate data visualisation,” in *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, ser. *UIST '17*. New York, NY, USA: Association for Computing Machinery, 2017, p. 71–83. [Online]. Available: <https://doi.org/10.1145/3126594.3126613>
- [68] P. Pirolli and S. Card, “The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis,” pp. 2–4, 2005. [Online]. Available: https://analysis.mitre.org/proceedings/Final_Papers_Files/206_Camera_Ready_Paper.pdf
- [69] E. Prather, A. S. Badr, B. Simões, and R. de Amicis, “A systematic literature review on dynamic cognitive augmentation through immersive reality: challenges and perspectives,” in *Defense + Commercial Sensing*, 2020.
- [70] J. G. Frederiksen, S. M. D. Sørensen, L. Konge, M. B. S. Svendsen, M. Nobel-Jørgensen, F. Bjerrum, and S. A. W. Andersen, “Cognitive load and performance in immersive virtual reality versus conventional virtual reality simulation training of laparoscopic surgery: a randomized trial,” *Surgical Endoscopy*, vol. 34, no. 3, pp. 1244–1252, Mar 2020. [Online]. Available: <https://doi.org/10.1007/s00464-019-06887-8>
- [71] C. M. de Melo, K. Kim, N. Norouzi, G. Bruder, and G. Welch, “Reducing cognitive load and improving warfighter problem solving with intelligent virtual assistants,” *Frontiers in Psychology*, vol. 11, 2020. [Online]. Available: <https://www.frontiersin.org/articles/10.3389/fpsyg.2020.554706>
- [72] P. Albus, A. Vogt, and T. Seufert, “Signaling in virtual reality influences learning outcome and cognitive load,” *Computers & Education*, vol. 166, p. 104154, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0360131521000312>

- [73] R. E. Mayer, *Cognitive Theory of Multimedia Learning*, ser. Cambridge Handbooks in Psychology. Cambridge University Press, 2005, p. 31–48.
- [74] Y.-C. Chen, Y.-S. Chang, and M.-J. Chuang, “Virtual reality application influences cognitive load-mediated creativity components and creative performance in engineering design,” *Journal of Computer Assisted Learning*, vol. 38, no. 1, pp. 6–18, 2022. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1111/jcal.12588>
- [75] L. J. Waguespack, “Visual metaphors for teaching programming concepts,” *SIGCSE Bull.*, vol. 21, no. 1, p. 141–145, feb 1989. [Online]. Available: <https://doi.org/10.1145/65294.71203>
- [76] G. Conti, E. Dean, M. Sinda, and B. Sangster, “Visual reverse engineering of binary and data files,” in *Visualization for Computer Security*, J. R. Goodall, G. Conti, and K.-L. Ma, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 1–17.
- [77] A. R. A. Grégio, A. O. C. Baruque, V. M. Afonso, D. S. F. Filho, P. L. de Geus, M. Jino, and R. D. C. dos Santos, “Interactive, visual-aided tools to analyze malware behavior,” in *Computational Science and Its Applications – ICCSA 2012*, B. Murgante, O. Gervasi, S. Misra, N. Nedjah, A. M. A. C. Rocha, D. Tanar, and B. O. Apduhan, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 302–313.
- [78] M. Wagner, F. Fischer, R. Luh, A. Haberson, A. Rind, D. A. Keim, and W. Aigner, “A Survey of Visualization Systems for Malware Analysis,” in *Eurographics Conference on Visualization (EuroVis) - STARs*, R. Borgo, F. Ganovelli, and I. Viola, Eds. The Eurographics Association, 2015.
- [79] A. Elliott, B. Peiris, and C. Parnin, “Virtual reality in software engineering: Affordances, applications, and challenges,” in *2015 IEEE/ACM 37th IEEE International Conference on Software Engineering*, vol. 2, 2015, pp. 547–550.
- [80] A. J. Ko, B. A. Myers, M. J. Coblentz, and H. H. Aung, “An exploratory study of how developers seek, relate, and collect relevant information during software maintenance

- tasks,” *IEEE Trans. Softw. Eng.*, vol. 32, no. 12, p. 971–987, dec 2006. [Online]. Available: <https://doi.org/10.1109/TSE.2006.116>
- [81] R. DeLine and K. Rowan, “Code canvas: zooming towards better development environments,” in *2010 ACM/IEEE 32nd International Conference on Software Engineering*, vol. 2, May 2010, pp. 207–210.
- [82] A. Bragdon, S. P. Reiss, R. Zeleznik, S. Karumuri, W. Cheung, J. Kaplan, C. Coleman, F. Adeputra, and J. J. LaViola, “Code bubbles: rethinking the user interface paradigm of integrated development environments,” in *2010 ACM/IEEE 32nd International Conference on Software Engineering*, vol. 1, May 2010, pp. 455–464.
- [83] J. Dominic, B. Tubre, J. Houser, C. Ritter, D. Kunkel, and P. Rodeghero, *Program Comprehension in Virtual Reality*. New York, NY, USA: Association for Computing Machinery, 2020, p. 391–395. [Online]. Available: <https://doi.org/10.1145/3387904.3389287>
- [84] S. G. Hart and L. E. Staveland, “Development of NASA-TLX (task load index): Results of empirical and theoretical research,” in *Human Mental Workload*, ser. Advances in Psychology, P. A. Hancock and N. Meshkati, Eds. North-Holland, 1988, vol. 52, pp. 139–183. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0166411508623869>
- [85] G. Lakoff, *Master Metaphor List*. University of California, 1994.
- [86] V. Averbukh, N. Averbukh, P. Vasev, I. Gvozdev, G. Levchuk, L. Melkozerov, and I. Mikhaylov, “Metaphors for software visualization systems based on virtual reality,” in *Augmented Reality, Virtual Reality, and Computer Graphics*, L. T. De Paolis and P. Bourdot, Eds. Cham: Springer International Publishing, 2019, pp. 60–70.
- [87] F. Fittkau, A. Krause, and W. Hasselbring, “Exploring software cities in virtual reality,” *2015 IEEE 3rd Working Conference on Software Visualization (VISSOFT)*, pp. 130–134, 2015.

- [88] R. Oberhauser and C. Lecon, “Gamified virtual reality for program code structure comprehension,” *International Journal of Virtual Reality*, vol. 17, no. 2, p. 79–88, Jan. 2017. [Online]. Available: <https://ijvr.eu/article/view/2894>
- [89] S. Romano, N. Capece, U. Erra, G. Scanniello, and M. Lanza, “On the use of virtual reality in software visualization: The case of the city metaphor,” *Information and Software Technology*, vol. 114, pp. 92–106, 2019. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0950584919301405>
- [90] N. Capece, U. Erra, S. Romano, and G. Scanniello, “Visualising a software system as a city through virtual reality,” in *Augmented Reality, Virtual Reality, and Computer Graphics*, L. T. De Paolis, P. Bourdot, and A. Mongelli, Eds. Cham: Springer International Publishing, 2017, pp. 319–327.
- [91] A. Hoff, L. Gerling, and C. Seidl, “Utilizing software architecture recovery to explore large-scale software systems in virtual reality,” in *2022 Working Conference on Software Visualization (VISSOFT)*, 2022, pp. 119–130.
- [92] A. Hoff, M. Lungu, C. Seidl, and M. Lanza, “Collaborative software exploration with multimedia note taking in virtual reality,” in *2024 IEEE/ACM 32nd International Conference on Program Comprehension (ICPC)*, 2024, pp. 346–357.
- [93] M. Weninger, L. Makor, and H. Mössenböck, “Memory cities: Visualizing heap memory evolution using the software city metaphor,” in *2020 Working Conference on Software Visualization (VISSOFT)*, 2020, pp. 110–121.
- [94] J. Auernhammer and B. Roth, “The origin and evolution of stanford university’s design thinking: From product design to design thinking in innovation management,” *Journal of Product Innovation Management*, vol. 38, no. 6, pp. 623–644, 2021. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1111/jpim.12594>
- [95] Design Council, “A study of the design process – the double diamond,” 2005, <https://www.designcouncil.org.uk/fileadmin/uploads/dc/>

Documents/ElevenLessons_Design_Council\%2520\%25282\%2529.pdf. [Online]. Available: https://www.designcouncil.org.uk/fileadmin/uploads/dc/Documents/ElevenLessons_Design_Council%2520%25282%2529.pdf

- [96] R. Clark, D. Feldon, J. J. G. Van Merrienboer, K. Yates, and S. Early, "Cognitive task analysis," in *Handbook of Research on Educational Communications and Technology*, J. M. Spector, M. D. Merrill, J. J. G. van Merrienboer, and M. P. Driscoll, Eds. New York: Routledge, 2008, ch. 43, pp. 577–593.
- [97] S. A. Mulder, "Cross-domain situational awareness in computing networks," *Sandia National Laboratories (SAND Report)*, 11 2014. [Online]. Available: <https://www.osti.gov/biblio/1494613>
- [98] The FLARE Team, "Capa, a tool to identify capabilities in programs and sandbox traces," <https://github.com/mandiant/capa>, 2024. [Online]. Available: <https://github.com/mandiant/capa>
- [99] Y. Shoshitaishvili, R. Wang, C. Salls, N. Stephens, M. Polino, A. Dutcher, J. Grosen, S. Feng, C. Hauser, C. Kruegel, and G. Vigna, "SoK: (State of) The Art of War: Offensive Techniques in Binary Analysis," in *IEEE Symposium on Security and Privacy*, 2016.
- [100] S. Weech, S. Kenny, and M. Barnett-Cowan, "Presence and cybersickness in virtual reality are negatively related: A review," *Frontiers in Psychology*, vol. 10, 2019. [Online]. Available: <https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2019.00158>
- [101] P. Healy and N. Nikolov, *Hierarchical Drawing Algorithms*. CRC Press, 08 2013, pp. 409–454.
- [102] M. Usuh, K. Arthur, M. C. Whitton, R. Bastos, A. Steed, M. Slater, and F. P. Brooks, "Walking ¿ walking-in-place ¿ flying, in virtual environments," in *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques*,

- ser. SIGGRAPH '99. USA: ACM Press/Addison-Wesley Publishing Co., 1999, p. 359–364. [Online]. Available: <https://doi.org/10.1145/311535.311589>
- [103] J. J. LaViola, “A discussion of cybersickness in virtual environments,” *SIGCHI Bull.*, vol. 32, no. 1, p. 47–56, Jan. 2000. [Online]. Available: <https://doi.org/10.1145/333329.333344>
- [104] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*. New York: Routledge, 1988.
- [105] S. Seabold and J. Perktold, “Statsmodels: Econometric and statistical modeling with python,” in *9th Python in Science Conference*, 2010.
- [106] J. Brooke, “SUS: A quick and dirty usability scale,” *Usability Eval. Ind.*, vol. 189, 11 1995.
- [107] B. G. Witmer, C. J. Jerome, and M. J. Singer, “The factor structure of the presence questionnaire,” *Presence*, vol. 14, no. 3, pp. 298–312, 2005.
- [108] T. Munzner, *Visualization Analysis and Design*. CRC Press, 2014.
- [109] C. Ware, *Information Visualization: Perception for Design*. Morgan Kaufmann, 2012.
- [110] E. R. Tufte, *The Visual Display of Quantitative Information*. Graphics Press, 1983.
- [111] Anthropic PBC, “Introduction: Model context protocol,” <https://modelcontextprotocol.io/>, 2025. [Online]. Available: <https://modelcontextprotocol.io/>
- [112] Canonical Ltd., “Ubuntu 20.04.6 lts (focal fossa),” <https://releases.ubuntu.com/focal/>, 2023. [Online]. Available: <https://releases.ubuntu.com/focal/>
- [113] OpenAI, Inc., “Models,” <https://platform.openai.com/docs/models>, 2025. [Online]. Available: <https://platform.openai.com/docs/models>

Appendices

Appendix A

Reverse Engineering Challenge source code listing

```
#include <stdio.h>

// ----- Part 1 Chain -----
int part1a()
{
    int input;
    printf("Enter integer 1: ");
    scanf("%d", &input);
    return (input == 5) ? input : 0;
}

int part1b()
{
    int a = part1a();
    return a * 2;    // 5 -> 10
}

int part1c()
{
    int a = part1a();
    int b = part1b();
    // Now depend on both part1a and part1b.
    // With a = 5 and b = 10, we want part1c to yield 4.
    return (b - a - 1); // 10 - 5 - 1 = 4
}

int part1d()
{
    int a = part1a();
    int b = part1b();
    int c = part1c();
    // Combine earlier results so that 10 + 4 + ? gives 9.
    // We'll use: (b - c + 3) so: 10 - 4 + 3 = 9.
}
```

```

    return (b - c + 3);
}

// ----- Part 2 Chain -----
int part2a()
{
    int input;
    printf("Enter integer 2: ");
    scanf("%d", &input);
    return (input == 7) ? input : 0;
}

int part2b()
{
    int a = part2a();
    return a * 3;    // 7 -> 21
}

int part2c()
{
    int a = part2a();
    int b = part2b();
    // With a = 7 and b = 21, we want part2c to yield 14.
    return (b - a); // 21 - 7 = 14
}

int part2d()
{
    int a = part2a();
    int c = part2c();
    // Combine a and c so that 14 + 7 + 2 = 23.
    return (c + a + 2);
}

// ----- Part 3 Chain -----
int part3a()
{
    int input;
    printf("Enter integer 3: ");
    scanf("%d", &input);
    return (input == 2) ? input : 0;
}

int part3b()
{
    int a = part3a();
    return a * 5;    // 2 -> 10
}

int part3c()
{
    int a = part3a();
    int b = part3b();
    // With a = 2 and b = 10, choose operations so that:
    // part3c = 10 - 2 - 4 = 4.
    return (b - a - 4);
}

```



```

int part3d()
{
    int b = part3b();
    int c = part3c();
    // With b = 10 and c = 4, compute:  $10 - 4 = 6$ .
    return (b - c);
}

// ----- Part 4 Chain -----
int part4a()
{
    int input;
    printf("Enter integer 4: ");
    scanf("%d", &input);
    return (input == 9) ? input : 0;
}

int part4b()
{
    int a = part4a();
    return a * 2; // 9 -> 18
}

int part4c()
{
    int a = part4a();
    int b = part4b();
    // With a = 9 and b = 18, choose:
    //  $\text{part4c} = 18 + 9 - 10 = 17$ .
    return (b + a - 10);
}

int part4d()
{
    int c = part4c();
    int c3 = part3c();
    //  $17 + 4 - 1 = 20$ 
    return (c + c3 - 1);
}

int part4e()
{
    int d = part4d();
    int a = part4a();
    // Use both d and a so that:  $(20 * 2) + 9 - 9 = 40$ .
    return (d * 2) + a - 9;
}

// ----- Part 5 Chain -----
int part5a()
{
    int input;
    printf("Enter integer 5: ");
    scanf("%d", &input);
    return (input == 3) ? input : 0;
}

```

```

int part5b()
{
    int a = part5a();
    return a * 3;    // 3 -> 9
}

int part5c()
{
    int a = part5a();
    int b = part5b();
    // With a = 3 and b = 9, combine them to get 11.
    return (b + a - 1);    // 9 + 3 - 1 = 11
}

int main()
{
    if ((part1d() + part3d() + part5c()) == ((part2d() + part4e()) / 3 + 5))
        printf("SUCCESS\n");
    else
        printf("FAILURE\n");

    return 0;
}

```

Listing A.1: Source for the RE Challenge Binary “v8”

Appendix B

User Study Playbook

User Study Playbook

Last updated 5/16/2025

Cheat Sheet

1. Before session: Prep the session spreadsheet on Box and make sure system works
2. Consent: Two forms to sign: Pre-Survey, Session
3. Screening and determine condition (VR-basic, VR-enhanced, Desktop/web)
4. Intro video for selected condition
5. Make sure Nexus is running
6. If VR condition: Check "feature_set" in config.ini (basic or enhanced)
7. If Web condition: open Ubuntu window; activate same Python environment as Nexus; cd programmingprojects/cogbre-web/cogbrew; npm start; go ahead and select collection (rechallenge) and binary (v8) for the user
8. Run the challenge: make sure they click the Q buttons and END button
9. Record question correctness and observations on the spreadsheet: notable behaviors and quotes
10. At end of session, if Web condition: click "Close session" button
11. Run post-survey for the selected condition
12. Copy the two session logs to the Box folder (only 1 log if Web condition)
13. Gift cards will be issued in batches within about a week after the session

Recruitment

Recruitment email with IRB stamp for reference:

https://drive.google.com/file/d/13z_TDGiMH9516x0fVgrf-2DFglzFv8u_/view?usp=drive_link

Recruitment email working doc:

https://docs.google.com/document/d/1cAA8aeCV-6vUbnVoku6z72Nbj4RLZz3UaOdgVhNpbZU/edit?usp=drive_link

Participant Timeslot Signup:

<https://doodle.com/sign-up-sheet/participate/4f2fb58b-332f-46a9-bf90-701e6705f221/select>

Participant Timeslot Signup management (probably only Dennis can access):

<https://doodle.com/sign-up-sheet/organize/4f2fb58b-332f-46a9-bf90-701e6705f221>

Investigator availability:

<https://www.when2meet.com/?29357050-ZSNaA>

Pre-Survey

Pre-survey will be taken by potential participants after viewing announcement email or otherwise finding out about the study.

Pre-survey will be taken online via Qualtrics.

Pre-survey requires its own consent, which is built into the pre-survey.

Qualtrics Pre-Survey link for participants (previously sent in email soliciting volunteers):

https://auburn.qualtrics.com/jfe/form/SV_3KnJwuEQ7boUnci

Pre-Survey Consent Form:

<https://auburn.box.com/s/op6ackh61j67qzi5l0g588avtvcvceok>

Pre-Survey IRB-stamped PDF for reference: <https://auburn.box.com/s/klimx8d74qy2qbasxghs7u3dpswxwwb9>

Pre-Survey edit view: https://auburn.yul1.qualtrics.com/survey-builder/SV_3KnJwuEQ7boUnci/edit

Before a Scheduled Session

Data collection setup:

- Go to Qualtrics and determine the unique ID assigned to the participant during the pre-survey
 - Go to https://auburn.yul1.qualtrics.com/responses/#/surveys/SV_3KnJwuEQ7boUnci
 - Log in to Qualtrics
 - Open Pre-survey project
 - Select “Data & Analysis” tab
 - Find Auburn user ID in the “Q2” column
 - **Participant ID** is in corresponding “Random ID” column
 - You will need to add this column if you haven’t previously

XM Pre-survey ▾

Survey Workflows Distributions Data & Analysis Results Reports

Data Text IQ Stats IQ Crosstabs IQ Weighting

Data Table Last Record Collected: 02/08/2025 7:53 PM MST Active Response Quality 100%

▼ Add Filter < 1 of 1 > 50 ▾

☐	Recorded Date	Q2 - Please enter your Auburn username, e.g., abc0003	Random ID	Q3 - Please enter your major field of study	soi
☐	Feb 8, 2025 7:53 PM	zzz1234	13003	None	
☐	Feb 8, 2025 7:48 PM		35875		

- Open up “_session data.xls” in the Box account using Excel
 - Location:
 - File in Excel: <https://auburn.app.box.com/integrations/officeonline/openOfficeOnline?fileId=1777341217760&sharedAccessCode=>
 - (if needed) Folder: <https://auburn.app.box.com/folder/306634350606>
 - (if needed) File: <https://auburn.box.com/s/gzwssccgs98o6dt6r9muq9vvx0c28xxj>
 - Go to “main data” tab
 - On the next open row, enter the **Date** and **Participant ID**
 - Make a new tab for observations using Participant ID and “obs.” For example, if participant is 12345, name the tab “12345 obs”
 - Move between the “main data” and new “obs” tabs as needed as you record data during the session
- REMINDERS:
 - **NO** identifying information (name, username, etc.) should go into a file stored on Box.
 - We **CANNOT** keep a list matching unique IDs to any identifying information. That is only stored in Qualtrics.

Cleaning:

- Ensure VR equipment is ready and sanitized. Use alcohol wipes to clean any surface that touches a participant such as the HMD face cushions, hand controllers, etc.

Start up CogBRE application components:

- Refer to “**Setting up VR for user study sessions**” in this document <https://docs.google.com/document/d/1UjEahcYpTgc48XrouTDOCUHUQr1-JckxSi2FT3A9-rs/edit?tab=t.0#heading=h.qsstis3zo0fk>

Running a Session

Description from IRB package item 13B: Individuals who choose to participate will first complete the online pre-survey hosted at Qualtrics or a similar service. This pre-survey will provide the first consent form, then collect contact information, demographic information, and a short self-rating of familiarity with computer architecture, compilers, reverse engineering of binary applications, and virtual reality. This step will take approximately 10 minutes. After completing the pre-survey, participants will sign up for a 60-minute time slot on SignUp Genius or a similar service. When the participant arrives for their scheduled session, they will be informed about the steps of the experiment and asked to sign the second consent form. Assuming they agree, the participant will then be directed to one of the three environments that we randomly assigned to them. If the participant is using the desktop environment, they will be provided access to a workstation with mouse and keyboard with software configured for the session; these items will be demonstrated to the participant. If the participant is using one of the two virtual reality (VR) environments, we will fit the HTC Vive headset to the participant and provide the two hand controllers. We will then familiarize them with the VR environment. These consent and familiarization steps are expected to take approximately 15 minutes. For any environment, we will next explain the reverse engineering challenge and let the participant formulate a solution on their own within that environment. This challenge is expected to take 10-15 minutes; we will cap this portion of the session at 30 minutes. As the participant works to solve a reverse engineering challenge in the desktop environment, actions such as executing command line actions will be logged. As the participant works to solve a reverse engineering challenge in the virtual reality environment, the participant's telemetry (position and orientation of head and hands) and actions will be logged. All logging is automated and transparent and logged data will be stored in the same secure folder as the electronic consent forms. When the participant finds the solution to the challenge, wants to stop, or hits the time limit, we will record the correctness of the solution and the time taken to find it in an electronic ledger stored in the secure folder. Finally, we will ask the participant to complete a post-survey on Qualtrics or a similar service using a workstation in the lab area. This step should take no more than 15 minutes, giving a total of 60 minutes for the participant. In between sessions, the VR equipment will be cleaned and sanitized.

Consent, Screening, Familiarization

Ensure you understand: *Participants will be allowed to take breaks, including temporarily removing the VR equipment, during the task. Participants will be told to immediately inform the researcher if they feel dizzy or uncomfortable with the VR equipment. If discomfort is expressed, the activity will cease immediately for that participant. They will be monitored by a researcher to ensure their well-being. If discomfort persists, appropriate recommendations for further assistance will be provided. In between sessions, the VR equipment will be cleaned and sanitized. To protect confidentiality, data from surveys will be coded, and participants' Auburn usernames will not be included. The results will only feature their survey responses and task performance alongside a code representing their identity. (Appendix D of IRB package)*

Consent -- Need to do **both** Pre-Survey and Session Consent forms!

- Use hard-copy consent forms (in case you need to print more):
 - PRE-SURVEY Consent Form PDF with IRB stamp:
<https://auburn.box.com/s/op6ackh61j67qzj5l0g588avtcvcueok>
 - SESSION Consent Form PDF with IRB stamp:
<https://auburn.box.com/s/y8btw2noikfl337jqpobi4afial1qyky>
- Have participant read consent forms and sign both if in agreement.
- File consent forms in designated folder locked in Dr. Mulder's office.

Screening and Familiarization:

- VR SCREENING (if applicable):
 - Give participant access to the screening questions on screen or as hardcopy.
 - Questions are in Appendix D (page 52) in the approved IRB package here:
https://drive.google.com/file/d/1k-3wqTHE_gv7GQWpn0ZABhvaqsmCGxVa/view?usp=drive_link
 - Copy of questions for convenience on next page

Review the questions below and answer them silently.

1. Do you have any cold or flu symptoms?
2. Do you have any open cuts or sores on your face or hands?
3. Do you have an empty stomach (have not eaten today)?
4. Do you have any other pre-existing medical conditions that would prevent effective use of the virtual reality equipment? (e.g., hand pain or deformity, strabismus (crossed eyes) or amblyopia (lazy eye), etc.)

If “yes” to any of the above, we unfortunately recommend you do not continue with this study.

5. Do you have any other pre-existing medical conditions that may impact the quality of your experience with the virtual reality equipment? (e.g., epileptic; pregnant; suffer from motion sickness or balance problems; susceptible to migraines; etc.)
6. Have you experienced eye strain?
7. Have you experienced stress induced anxiety?

If “yes” to any of the above, proceed with caution and let the facilitator know if you are experiencing any symptoms during the virtual reality experience.

- If participant doesn't pass screening, don't continue with VR condition
- Determine which condition the participant will experience
- Briefly describe what the participant will experience during the experiment based on the condition they are assigned (desktop, low-affordance VR, or high-affordance VR).
- If using VR, show VR equipment.

Show Intro video

Show intro video for the experimental condition to which the participant is assigned:
WEB, VR BASIC, or VR ENHANCED

Videos are in BOX folder linked below under “_intro videos”. Probably need to copy to your local computer.

<https://auburn.app.box.com/folder/306634350606>

Get System Ready to Perform Challenge

Ensure Nexus is running regardless of condition (Desktop or VR)

Config.ini is located in the root of the VR project. (cogbre-vr)

Condition 1: Desktop

- Start web app:
 - Open Terminal
 - go to cogbre-web/cogbrewweb folder
 - > npm start
 - Browser should open automatically
 - If not, open one and go to localhost:3000
- Go ahead and select collection (rechallenge) and binary (v8) for the user

Condition 2: Basic VR

- Ensure in config.ini that “feature_set = basic_VR_user_study”
- In Unity editor, run cogbre-vr
- Help participant put on HMD and hold controllers
 - Make necessary physical adjustments and **IPD**
 - Re-center VR (if necessary) at the center of the physical play area. Position matters but orientation doesn't really matter; we just need to be sure the participant has room to safely work.
- Help user select the correct Collection (rechallenge) and Binary file (v8)

Condition 3: Enhanced VR

- Ensure in config.ini that “feature_set = enhanced_VR_user_study”
- In Unity editor, run cogbre-vr
- Help participant put on HMD and hold controllers
 - Make necessary physical adjustments and **IPD**
 - Re-center VR (if necessary) at the center of the physical play area. Position matters but orientation doesn't really matter; we just need to be sure the participant has room to safely work.

- Help user select the correct Collection (rechallenge) and Binary file (v8)

Perform Reverse Engineering Challenge

- **Ready the spreadsheet**
 - **General approach:** Assist user with questions about how to use the tool and with technical issues (including restarting software if needed) -- but don't assist with solving challenge.
 - "Think aloud": Encourage the participant to "think aloud" as they are working. This is NOT a requirement and they can decline. During the session, make notes and observations in the "##### obs" tab you created earlier.
 - **Record** session start time on spreadsheet just before asking first question
 - Ask each question in order.
 - When user gives answer, **record the answer correctness and time** on spreadsheet.
 - If user gives correct answer, tell them it was correct.
 - If not, tell them the correct answer.
 - Session ends when any of these occur:
 - All questions are answered
 - Hit time limit of 30 minutes
 - Participant wants to quit
 - **Record** the time the session ends in the spreadsheet
 - If using Web app, **click "Close Session"** at top of page or else we'll lose data!
 - If using VR, stop the VR playback in Unity Editor
 - This is necessary for Nexus to properly close the log files
 - If using VR, remove equipment
 - Provide recovery time before moving to post-survey (if needed)
- ==> Ensure participant is feeling well. If not, follow University protocol for illness.

QUESTIONS TO ASK PARTICIPANT

⇒ ask participant to click “Q1” on the question control box

Question 1: “In this program, a user's input is considered correct when a function that asks for an input returns the user's input value instead of zero. What specific input does the function part1a require to return a nonzero value?”

Answer: 5

⇒ ask participant to click “Q2” on the question control box

Question 2: “What final value is returned by part1d when the input to part1a is correct?”

Answer: 9

⇒ ask participant to click “Q3” on the question control box

Question 3: “What final value is returned by part3d when the input to part3a is correct?”

Answer: 6

⇒ ask participant to click “Q4” on the question control box

Question 4: “What final value is returned by part4e when the required inputs are correct?”

Answer: 40

⇒ ask participant to click “Q5” on the question control box

Question 5: “How many times is the function part1a called if the part1d function fully executes (you examined this chain of functions in question 2)?”

Answer: 4

⇒ ask participant to click “END” on the question control box

Wrap up: Post-Survey, Debrief, etc.

Post-survey depends on experimental condition for the participant:

WEB:

- **Use shortcut** for Web Post-Survey in “CogBRE User Study Shortcuts” on reverse3 desktop
- **Or use direct Qualtrics link for participant:** https://auburn.qualtrics.com/jfe/form/SV_0pjW94OJq1gMxDw
- Consists of TLX and SUS; IRB-stamped PDF for reference:
https://drive.google.com/file/d/1XZOkca1T1rk_aUaEZbcHlISnkgwU2ZKI/view?usp=drive_link

VR BASIC:

- **Use shortcut** for VR Basic Post-Survey in “CogBRE User Study Shortcuts” on reverse3 desktop
- **Or use direct Qualtrics link for participant:** https://auburn.qualtrics.com/jfe/form/SV_7R4Q1JzynDVUbHM
- Consists of TLX, SUS, Presence; IRB-stamped PDF for reference:
https://drive.google.com/file/d/1MJ8e-Jyxvql8HYF7febnFdt3NnD4qDcE/view?usp=drive_link

VR ENHANCED:

- **Use shortcut** for VR Enhanced Post-Survey in “CogBRE User Study Shortcuts” on reverse3 desktop
- **Or use direct Qualtrics link for participant:** https://auburn.qualtrics.com/jfe/form/SV_5mtQkcRJ2canLX8
- Consists of TLX, SUS, Presence, Likert questions about VR affordances; IRB-stamped PDF for reference:
https://drive.google.com/file/d/1ja6WDHnjWULnqkBVnup3mvV5jgZZqyfb/view?usp=drive_link

De-brief: The participant performed a basic reverse engineering challenge in one of the above conditions. We monitored their actions, which we will analyze along with other participant data to find correlations between how they used the environment and their performance, if any.

ENSURE THE PARTICIPANT KNOWS NOT TO SHARE DETAILS OF THE QUESTIONS AND ANSWERS TO OTHER POTENTIAL PARTICIPANTS!

Ask participant for the email to which the Gift Card should be sent.

Thank participant for their time.

Saving Session Data

- Find session log file(s):
 - Look in the folder cogbre-nexus/sessions
 - You can open this folder in Windows Explorer using:
\\wsl.localhost\Ubuntu\home\reverse_3\Programming_Projects\cogbre_nexus\sessions

- You should see 2 new files for VR condition (events and telemetry) or 1 new file for Desktop condition (events)
 - The filename(s) contain the unique session ID for the Nexus session, e.g., “1b39bf5839204a2fbe771d721e867081”
- Copy all session log files into the same Box folder that holds the spreadsheet
 - **NOTE:** If the Desktop web app was refreshed during the session, then there may be multiple log files. Ensure you copy all files from the session to Box.
- Enter the unique session ID into the **nexus_guid** field in the spreadsheet

Session Management

Simplymeet admin link: https://secure.simplymeet.me/event_types/#/214857/edit

Doodle admin link (NO LONGER IN USE AFTER 3/25):

<https://doodle.com/sign-up-sheet/organize/4f2fb58b-332f-46a9-bf90-701e6705f221>

Simplymeet welcome message

Thank you for volunteering for the user study “Impacts of Immersive Virtual Reality on the Binary Reverse Engineering Process!”

Your session will take place in the Program Understanding Lab, which is in Room 2119 in the Shelby Center.

When you arrive, either Dennis Brown or Kevan Baker will be in the lab to run your session.

We look forward to your timely arrival and appreciate your support!

Please sign up for a time slot below.

You will then receive email verification from this system, and we will follow up later with a calendar invitation.

Calendar Meeting Text

Thank you for signing up for the user study “Impacts of Immersive Virtual Reality on the Binary Reverse Engineering Process!”

Your session will take place in the Program Understanding Lab, which is in Room 2119 in the Shelby Center.

When you arrive, either Dennis Brown or Kevan Baker will be in the lab to run your session.

We look forward to your timely arrival and appreciate your support!

If you have any questions or concerns, please let us know.

Reminder for unscheduled participants

Good afternoon,

Thank you for completing the pre-survey for my study!

I did not see that you have signed up for a session time (please correct me if I'm wrong).

As a reminder, you can sign up for a time here:

<https://app.simplymeet.me/dbrown/pul-60-1742654987953>

If you have issues with that site, or need to schedule a time not listed, please just let me know!

Thank you,

Dennis Brown

Appendix C

Institutional Review Board Package

Revised 07/12/2022

AUBURN UNIVERSITY INSTITUTIONAL REVIEW BOARD for RESEARCH INVOLVING HUMAN SUBJECTS**PROTOCOL REVIEW FORM
FULL BOARD or EXPEDITED REVIEW**For assistance, contact: **The Office of Research Compliance (ORC)**Phone: **334-844-5966** E-Mail: IRBAdmin@auburn.edu Web Address: <http://www.auburn.edu/research/vpr/ohs>**Submit completed form and supporting materials as one PDF through the [IRB Submission Page](#)****Handwritten forms are not accepted. Where links are found hold down the control button (Ctrl) then click the link.**

1. Proposed Start Date of Study: 9/2/2024 Today's Date: **July 15, 2024**
 Submission Status (Check One): ☐ New ☒ Revisions (to address IRB Review Comments)
 Proposed Review Category (Check One): ☐ Full Board (greater than minimal risk) ☒ Expedited
 If Expedited, Indicate Category(ies) (([Link to Expedited Category Review Sheet](#)) 7

2. Project Title: Impacts of Immersive Virtual Reality on the Binary Reverse Engineering Process

3. Principal Investigator (PI): Dennis G. Brown Degree(s): MS & BA CompSci
 Rank/Title: Graduate Student Department/School: Computer Science and Software Engineering
 Role/responsibilities in this project: Research, conceptualization, experimental design, software development, informed consent, risk and safety monitoring, conducting sessions, data collection, data analysis, report writing.
 Preferred Phone Number: 478-550-0247 AU Email: dgb0028@auburn.edu

Faculty Advisor Principal Investigator (if applicable): Dr. Samuel Mulder
 Rank/Title: Associate Professor Department/School: Computer Science and Software Engineering
 Role/responsibilities in this project: Supervise the development, conducting, and analysis of this research project
 Preferred Phone Number: 334-844-6321 AU Email: szm0211@auburn.edu

Department Head: Dr. Hari Narayanan Department/School: Computer Science and Software Engineering
 Preferred Phone Number: 334-844-6312 AU Email: naraynh@auburn.edu
 Role/responsibilities in this project: Review and approve the project

4. Funding Support: ☐ N/A ☒ Internal External Agency: n/a Pending ☐ Received ☒
 For federal funding, list funding agency and grant number (if available): n/a

5. a) List any contractors, sub-contractors, and other entities associated with this project: none
 b) List any other AU IRB approved protocols associated with this study and describe the association: none
 c) List any other institutions associated with this study and submit a copy of their IRB approval(s): none

Protocol Packet Checklist**Check all applicable boxes. A completed checklist is required.**

- ☒ **Protocol Review Form** (All required signatures included and all sections completed)
 (Examples of appended documents are found on the website: <https://cws.auburn.edu/OVPR/pm/compliance/irb/sampledocs>)
- ☒ **CITI Training Certificates** for key personnel
- ☒ **Consent Form or Information Letter** and any releases (audio, video or photo) that participants will review and/or sign
- ☒ **Appendix A** "Reference List"
- ☒ **Appendix B** if e-mails, flyers, advertisements, social media posts, generalized announcements or scripts, etc., will be used to recruit participants.
- ☒ **Appendix C** if data collection sheets, surveys, tests, other recording instruments, interview scripts, etc. will be used for data collection. Attach documents in the order they are listed in item 13c. **Continued on Page 2**
- ☒ **Appendix D** if they study will use a debriefing form or will include emergency plans/ procedures and medical referral lists. (A referral list may be attached to the consent document.)

The Auburn University Institutional
 Review Board has approved this
 Document for use from
 07/23/2024 to _____
 Protocol # 24-668 EP 2407

Revised 07/12/2022

- ☐ **Appendix E** If research is being conducted at sites other than Auburn University or in cooperation with other entities. A **permission letter** from the site/ program director must be included indicating their cooperation or involvement in the project. NOTE: If the proposed research is a multi-site project, involving investigators or participants at other academic institutions, hospitals or private research organizations, a letter of **IRB approval** from each entity is required prior to initiating the project.
- ☐ **Appendix F** Written evidence of approval by the host country, local IRB or institutions if research is conducted outside the United States

6. General Research Project Characteristics

6A. Research Methodology	
Check all descriptions that best apply to the research methodology.	
Data Source(s): ☒ New Data ☐ Existing Data	Will recorded data directly or indirectly identify participants? ☒ Yes ☐ No
Data collection will involve the use of: ☐ Educational Tests (cognitive diagnostic, aptitude, etc.) ☒ Interview ☒ Observation ☐ Locations or Tracking Measures ☒ Physical / Physiological Measures or Specimens ☒ Surveys / Questionnaires ☐ Other: Click or tap here to enter text. ☒ Internet / Electronic ☐ Audio ☐ Video ☐ Photos ☐ Digital Images ☐ Private records or files	
6B. Participant Information	6C. Risks to Participants
Check all descriptors that apply to the TARGET population. (link to definition of target population) ☒ Males ☒ Females ☒ AU students Vulnerable Populations ☐ Pregnant Women/Fetuses ☐ Prisoners ☐ Institutionalized ☐ Children and / or Adolescents (under age 18 in AL; if minor participants, at least 2 adults must be present during all research procedures that include the minors) Persons with: ☐ Economic Disadvantages ☐ Physical Disabilities ☐ Educational Disadvantages ☐ Intellectual Disabilities Will participants be compensated? ☒ Yes ☐ No	Identify all risks participants might encounter in this research. ☒ Breach of Confidentiality* ☐ Deception ☐ Psychological ☐ None ☐ Other (COVID-19, other medical): n/a ☐ Coercion ☒ Physical ☐ Social *Note that if the investigator is using or accessing confidential or identifiable data, reach of confidentiality is always a risk.
6D. Corresponding Approval/ Oversight	
- Does the study include participant exposure to radiation? ☐ Yes ☒ No If yes indicate: ☐ DEXA ☐ PQCT ☐ Other - Is IBC Approval required for this study? ☐ Yes ☒ No If yes, BUA # Click or tap here to enter text. Expiration Date Click or tap to enter a date. - Is IACUC Approval required for this study? ☐ Yes ☒ No If yes, PRN # Click or tap here to enter text. Expiration Date Click or tap to enter a date. - Does this study involve the Auburn University MRI Center? ☐ Yes ☒ No Which MRI(s) will be used for this project? (Check all that apply) ☐ 3T ☐ 7T Does any portion of this project require review by the MRI Safety Advisory Council?	

Continued on Page 3

Revised 07/12/2022

☐ Yes☐ No

Signature of one MRI Center Representative: _____

Required for all projects involving the AU MRI Center

Appropriate MRI Center Representatives:

Dr. Thomas S. Denney, Director AU MRI Center

Dr. Ron Beyers, MR Safety Officer

7. Project Assurances**7A. Principal Investigator's Assurances**

1. I certify that all information provided in this application is complete and correct.
2. I understand that, as Principal Investigator, I have ultimate responsibility for the conduct of this study, the ethical performance of this project, the protection of the rights and welfare of human subjects, and strict adherence to any stipulations imposed by the Auburn University IRB.
3. I certify that all individuals involved with the conduct of this project are qualified to carry out their specified roles and responsibilities and are in compliance with Auburn University policies regarding the collection and analysis of the research data.
4. I agree to comply with all Auburn policies and procedures, as well as with all applicable federal, state, and local laws regarding the protection of human subjects, including, but not limited to the following:
 - a. Conducting the project by qualified personnel according to the approved protocol
 - b. Implementing no changes in the approved protocol or consent form without prior approval from the Office of Research Compliance
 - c. Obtaining the legally effective informed consent from each participant or their legally responsible representative prior to their participation in this project using only the currently approved, stamped consent form
 - d. Promptly reporting significant adverse events and / or effects to the Office of Research Compliance in writing within 5 working days of the occurrence.
5. If I will be unavailable to direct this research personally, I will arrange for a co-investigator to assume direct responsibility in my absence. This person has not been named as co-investigator in this application, or I will advise ORC, by letter, in advance of such arrangements.
6. I agree to conduct this study only during the period approved by the Auburn University IRB.
7. I will prepare and submit a renewal request and supply all supporting documents to the Office of Research Compliance before the approval period has expired if it is necessary to continue the research project beyond the time period approved by the Auburn University IRB.
8. I will prepare and submit a final report upon completion of this research project.

My signature indicates I have read, understand and agree to conduct this research project in accordance with the assurances listed above.

Dennis G. Brown

Principal Investigator Name

Dennis G. Brown

Principal Investigator Signature

7/1/24

Date

7B. Faculty Advisor / Sponsor's Assurances

1. I have read the protocol submitted for this project for content, clarity, and methodology.
2. By my signature as faculty advisor / sponsor on this research application, I certify that the student or guest investigator is knowledgeable about the regulations and policies governing research with human subjects and has sufficient training and experience to conduct this particular study in accord with the approved protocol.
3. I agree to meet with the investigator on a regular basis to monitor study progress. Should problems arise during the course of the study, I agree to be available, personally, to supervise the investigator in solving them.
4. I assure that the investigator will promptly report significant incidents and / or adverse events and / or effects to the ORC in writing within 5 working days of the occurrence.
5. If I will be unavailable, I will arrange for an alternate faculty sponsor to assume responsibility during my absence, and I will advise the ORC by letter of such arrangements. If the investigator is unable to fulfill requirements for submission of renewals, modifications or the final report, I will assume that responsibility.

Dr. Samuel Mulder

Faculty Advisor / Sponsor Name

Samuel Mulder

Faculty Advisor Signature

7/1/24

Date

Continued on Page 4

Revised 07/12/2022

7C. Department Head's Assurance

By my signature as department head, I certify that I will cooperate with the administration in the application and enforcement of all Auburn University policies and procedures, as well as all applicable federal, state, and local laws regarding the protection and ethical treatment of human participants by researchers in my department

Dr. Hari Narayanan

Department Head Name

H. Narayanan
Department Head Signature

7/1/24
Date

8. Project Overview:

8A. A summary of relevant research findings leading to this research proposal:

(Cite source; include a "Reference List" as [Appendix A.](#))

Binary reverse engineering (RE), also known as program understanding or program comprehension, is a difficult analytical task, especially crucial to cybersecurity [1] [2]. Our work explores how immersive virtual reality (VR), leveraging principles of embodied and external cognition, can improve the performance of RE practitioners. While we have found no prior work specifically applying VR to binary RE, we found related work in source code RE / comprehension. Elliott et al. [3] explored the use of affordances of VR such as spatial cognition, presence, and motion in navigating and comprehending source code. Dominic et al. [4] performed a user study comparing source code comprehension between a traditional 2D desktop interface and VR, finding no advantage to VR; however, their VR environment simply replicated the 2D desktop and was not tailored to take advantage of specific VR affordances. Romano et al. [5] performed an experiment with 42 participants (3 conditions, between-subjects design) demonstrating significantly higher source code comprehension using their VR interface, based on navigating a city metaphor representing the code, compared to a traditional integrated development environment (IDE). Hoff et al. [6] performed a similar experiment with 54 participants (3 conditions, between-subjects design) focused on understanding software architecture, demonstrating similar or better performance of participants using their VR interface, based on navigating a solar system metaphor, compared to a traditional IDE. These findings show that VR has improved source code RE under certain circumstances, and justifies running a study to measure the impacts of VR on binary RE as we propose here.

8B. A brief summary/abstract of the study methodology, including design, population, and variables of interest.

(350 word maximum, in language understandable to someone who is not familiar with your area of study. Note this summary/abstract can be used to prepare the concise summary in the consent document.):

This study seeks to measure the effectiveness of different configurations of tools designed to facilitate the process of reverse engineering a binary program. These configurations will include a traditional computer desktop environment and two immersive virtual reality (VR) environments. Participants will execute one program understanding task (reverse engineering challenge) in one of these three environments (between-subjects design). The independent variable is the environment in use. Dependent variables are task performance metrics (e.g., accuracy and speed), system usability (survey), cognitive load (survey), immersive (survey), and feature rating (survey); the system will also log their actions in the environments. In the traditional approach, they will use reverse engineering tools in a common desktop windowed environment controlled by keyboard and mouse. In the VR environments, they will wear an HTC Vive Pro 2 VR headset and use two HTC VR controllers. The study will occur in the Shelby Center on the Auburn campus. The population will be Auburn computer science and engineering students with an understanding of compilers and binary applications. We estimate requiring at least 54 complete and valid participant sessions. We will offer a \$20 gift card to 70 participants as an incentive. Each participant will attempt to solve a single reverse engineering challenge in one of the three environments. The reverse engineering challenge will be targeted to take up to 30 minutes to solve; participants can stop at any time. Complete sessions, from consent to performing the task to answering a survey, should take no longer than 60 minutes per participant.

9. Purpose

9A. State the purpose of the study and all research questions or aims. (Include a sentence that begins, "The purpose of this study is...")

The purpose of this study is to discover quantifiable measures of effectiveness of using immersive virtual reality (VR) affordances in the process of performing binary reverse engineering tasks. This study is predicated on theories of cognition in analytics tasks and effects of immersive VR. Our hypotheses include: (1) Participants will have better task

Revised 07/12/2022

completion metrics with certain levels of VR affordances than other levels. (2) Cognitive load will vary by level of VR affordances. (3) Participants will rate certain VR affordances (e.g., user organization, spatial memory, signalling) higher than others in terms of usefulness. (4) Participants who have better performance in the immersive environments will have used more of the VR space.

9B. Describe how results of this study will be used? (e.g., presentation? publication? thesis? dissertation?)

The results of this study will be written up in a dissertation and associated local presentations (e.g., dissertation defense). The results will also be included in papers submitted to international conferences and journals, which may have associated presentations.

10. Key Personnel. Describe responsibilities as specifically as possible. Include information on research training or certifications related to this project. **To determine key personnel see decision tree at <https://cws.auburn.edu/OVPR/pm/compliance/irb/training>. Submit a copy of CITI training documentation for all key personnel.** (For additional personnel, add lines as needed).

To determine Auburn University HIPAA – covered entities click link to [HIPAA Policy](#).

If any key personnel have a formal association with institutions/entities involved in the study (for example is an employee or supervisor at the site research will occur), describe that affiliation. For all non-AU affiliated key personnel, submit a copy of their IRB approval.

Principal Investigator: Dennis Brown
Email Address: dgb0028@auburn.edu

Rank/Title: Graduate Student

Degree(s): MS, BA CompSci

Dept / Affiliation: Computer Science and Software Engineering

HIPAA Covered Entity? Yes ☐ No ☒

Roles / Responsibilities: Research, conceptualization, experimental design, software development, informed consent, risk and safety monitoring, conducting sessions, data collection, data analysis, report writing.

- AU affiliated? ☒ Yes ☐ No If no, name of home institution: n/a

- Plan for IRB approval for non-AU affiliated personnel? n/a

- Do you have any known competing financial interests, personal relationships, or other interests that could have influence or appear to have influence on the work conducted in this project? ☐ Yes ☒ No

- If yes, briefly describe the potential or real conflict of interest: n/a

- Completed required CITI training? ☒ Yes ☐ No If NO, complete the appropriate [CITI basic course](#) and update the revised Exempt Application form.

- If YES, choose course(s) the researcher has completed: Human Sciences Basic Course 11/5/2024

Choose a course Expiration Date

Individual: Dr. Samuel Mulder

Email Address: szm0211@auburn.edu

Rank/Title: Associate Research Professor

Degree(s): PhD, MS, BS CompSci

Dept. / Affiliation: Computer Science and Software Engineering

HIPAA Covered Entity? Yes ☐ No ☒

Roles / Responsibilities: Supervise the development, conducting, and analysis of this research project

- AU affiliated? ☒ Yes ☐ No If no, name of home institution: n/a

- Plan for IRB approval for non-AU affiliated personnel? n/a

- Do you have any known competing financial interests, personal relationships, or other interests that could have influence or appear to have influence on the work conducted in this project? ☐ Yes ☒ No

- If yes, briefly describe the potential or real conflict of interest: n/a

- Completed required CITI training? ☐ Yes ☐ No If NO, complete the appropriate [CITI basic course](#) and update the revised Exempt Application form.

- If YES, choose course(s) the researcher has completed: Human Sciences Basic Course 7/1/2027

Choose a course Expiration Date

Revised 07/12/2022

Individual: Kevan Baker

Email Address: kzb0154@auburn.edu

Dept. / Affiliation: Computer Science and Software Engineering

Roles / Responsibilities: Software development, informed consent, risk and safety monitoring, conducting sessions, data collection, data analysis, report writing

- AU affiliated? ☒ Yes ☐ No If no, name of home institution: n/a
- Plan for IRB approval for non-AU affiliated personnel? n/a
- Do you have any known competing financial interests, personal relationships, or other interests that could have influence or appear to have influence on the work conducted in this project? ☐ Yes ☒ No
- If yes, briefly describe the potential or real conflict of interest: n/a
- Completed required CITI training? ☒ Yes ☐ No If NO, complete the appropriate [CITI basic course](#) and update the revised Exempt Application form.
- If YES, choose course(s) the researcher has completed: Human Sciences Basic Course 6/24/2027

[Choose a course](#) [Expiration Date](#)

Individual: Julian Bauer

Email Address: jcb0209@auburn.edu

Dept. / Affiliation: Computer Science and Software Engineering

Roles / Responsibilities: Software development, informed consent, risk and safety monitoring, conducting sessions, data collection, data analysis, report writing

- AU affiliated? ☒ Yes ☐ No If no, name of home institution: n/a
- Plan for IRB approval for non-AU affiliated personnel? n/a
- Do you have any known competing financial interests, personal relationships, or other interests that could have influence or appear to have influence on the work conducted in this project? ☐ Yes ☒ No
- If yes, briefly describe the potential or real conflict of interest: n/a
- Completed required CITI training? ☒ Yes ☐ No If NO, complete the appropriate [CITI basic course](#) and update the revised Exempt Application form.
- If YES, choose course(s) the researcher has completed: Human Sciences Basic Course 6/26/2027

[Choose a course](#) [Expiration Date](#)

Individual:

Email Address:

Dept. / Affiliation: Computer Science and Software Engineering

Roles / Responsibilities: [Click or tap here to enter text.](#)

- AU affiliated? ☐ Yes ☐ No If no, name of home institution: n/a
- Plan for IRB approval for non-AU affiliated personnel? n/a
- Do you have any known competing financial interests, personal relationships, or other interests that could have influence or appear to have influence on the work conducted in this project? ☐ Yes ☒ No
- If yes, briefly describe the potential or real conflict of interest: n/a
- Completed required CITI training? ☐ Yes ☐ No If NO, complete the appropriate [CITI basic course](#) and update the revised Exempt Application form.
- If YES, choose course(s) the researcher has completed: [Choose a course](#) [Expiration Date](#)

[Choose a course](#) [Expiration Date](#)

11. Location of research.

11A. List all locations where data collection will occur. If applicable, attach permission letters as Appendix

E. (School systems, organizations, businesses, buildings and room numbers, servers for web surveys, etc.) **Be as specific as possible.**

Revised 07/12/2022

(See sample letters at <https://cws.auburn.edu/OVPR/pm/compliance/irb/sampledocs>)

The study will occur in the Program Understanding Lab space in the Shelby Center on the Auburn campus.

11B. Will study data be stored within a HIPAA covered facility? Yes ☐ No ☒

If yes, which facility(ies) (To determine AU HIPAA covered entities, go to VII of the [HIPAA Hybrid Entity Policy](#)):
n/a

12. Participants (If minor participants, at least 2 adults must be present during all research procedures that include the minors.)

12A. Describe the targeted/ intended participant population for the study. Include the anticipated number of participants and inclusion and exclusion criteria and the procedures to ensure more than 1 adult is present during all research procedures which include the minor.

☐ Check here if existing data will be used and describe the population from whom data was collected including the number of data files.

☐ Check here if permission to access existing data is required and submit a copy of the agreement to access.

We will recruit from the population of Auburn computer science and engineering students with an understanding of compilers and binary applications, who will typically be from the upper classes or graduate programs. Minors will be excluded. We estimate requiring at least 54 complete and valid participant sessions, so we will target 70 participants to account for dropouts (results from all valid sessions will be included in the analysis). Participants will not be eligible to participate in the experiment if they have contraindications to the use of the VR equipment, including cold, flu, headache, migraines, earache, severe deformity or pain in the hand, pregnancy, pre-existing binocular vision abnormalities (strabismus (crossed eyes) or amblyopia (lazy eye)), or other serious medical condition.

12B. Describe, step-by-step in lay language all procedures to recruit participants. Include in [Appendix B](#) a copy of all e-mails, flyers, advertisements, recruiting scripts, invitations, etc., that will be used to invite people to participate. (See sample documents at <https://cws.auburn.edu/OVPR/pm/compliance/irb/sampledocs>)

Recruiting will occur primarily via email containing the language included in Appendix B. The email will be sent to engineering student organizations and students via faculty. The email will direct interested individuals to complete a pre-survey (also in Appendix B) on Qualtrics or a similar service, and then to sign up for a time slot using Signup Genius or a similar service.

12C. Minimum number of participants required to validate the study? 54

Number of participants expected to enroll? 70

Provide the rationale for the number of participants. The study uses a between-subjects design. The expected magnitude of the differences we expect to detect is medium-to-high, estimated at $d = 0.7$ based on Cohen's "Statistical power analysis for the behavioral sciences," similar to other VR studies. We will analyze the results using Analysis of variance (ANOVA). Using statistical power = 0.8, significance level = 0.05, and $d = 0.7$ in a simple power analysis (executed via the Python module Statsmodels), our experiment needs 18 samples per condition. With 3 conditions we require 54 valid sessions, so we will recruit 70. The two similar studies referenced in 8A above had numbers of participants in this same range.

Is there a limit to the number of participants that will be included in the study?

☐ No ☒ Yes, the number is 70

12D. Describe the process to compensate, amount and method of compensation and/or incentives for participants. [AU Procurement and Business Services \(PBS\) policies](#)
(benefits to participants are NOT compensation)

If participants will not be compensated, check here: ☐

Indicate the amount of compensation per procedure and in total: \$20 gift card per participant; \$1400 total.

Revised 07/12/2022

- Indicate the type of compensation: ☒ Monetary ☐ Incentives
- ☐ Raffle or Drawing incentive (Include the chances of winning.)
 - ☐ Extra Credit (State the value)
 - ☐ Other

Describe how compensation will be distributed (USPS, email, etc.): Compensation will be distributed via email to those selected.

13. Project Design & Methods

13A. Describe, step-by-step, all procedures and methods that will be used to consent participants. If a waiver is being requested, indicate the waiver, and describe how the study meets the criteria for the waiver. If minors will be enrolled describe the process to obtain parental/ legally authorized guardian permission.

- ☐ Waiver of Consent (including using existing data)
- ☐ Waiver of Documentation of Consent (use of Information Letter)
- ☐ Waiver of Parental Permission (for college students 18 years or younger)

We will use a two-part consent process. Part 1: During the recruitment process, interested individuals will complete a pre-survey online using Qualtrics or a similar service, and then choose a time slot via SignUp Genius or a similar service. During this sign up process, individuals will be asked to consent to providing their Auburn ID, first and last names, and school email address for purposes of connecting their pre-survey responses to their session results and preparing the incentive electronic gift card. Part 2: When a participant arrives for their scheduled session, they will be asked to consent on a paper form to participate in the session, have their actions observed and logged by the system, and to complete a post-task survey. The pre-survey consent will remain on the approved online survey tool. The session consent paper forms will be stored in Dr. Mulder's office. Participants will be provided copies of the consent forms.

13B. In lay language, understandable by someone not familiar with the area of study, describe the complete research design and methods that will be used to address the purpose. Include a clear description of who, when, where and how data will be collected. Include specific information about participants' time and effort.

Individuals who choose to participate will first complete the online pre-survey hosted at Qualtrics or a similar service. This pre-survey will provide the first consent form, then collect contact information, demographic information, and a short self-rating of familiarity with computer architecture, compilers, reverse engineering of binary applications, and virtual reality. This step will take approximately 10 minutes. After completing the pre-survey, participants will sign up for a 60-minute time slot on SignUp Genius or a similar service. When the participant arrives for their scheduled session, they will be informed about the steps of the experiment and asked to sign the second consent form. Assuming they agree, the participant will then be directed to one of the three environments that we randomly assigned to them. If the participant is using the desktop environment, they will be provided access to a workstation with mouse and keyboard with software configured for the session; these items will be demonstrated to the participant. If the participant is using one of the two virtual reality (VR) environments, we will fit the HTC Vive headset to the participant and provide the two hand controllers. We will then familiarize them with the VR environment. These consent and familiarization steps are expected to take approximately 15 minutes. For any environment, we will next explain the reverse engineering challenge and let the participant formulate a solution on their own within that environment. This challenge is expected to take 10-15 minutes; we will cap this portion of the session at 30 minutes. As the participant works to solve a reverse engineering challenge in the desktop environment, actions such as executing command line actions will be logged. As the participant works to solve a reverse engineering challenge in the virtual reality environment, the participant's telemetry (position and orientation of head and hands) and actions will be logged. All logging is automated and transparent and logged data will be stored in the same secure folder as the electronic consent forms. When the participant finds the solution to the challenge, wants to stop, or hits the time limit, we will record the correctness of the solution and the time taken to find it in an electronic ledger stored in the secure folder. Finally, we will ask the participant to complete a post-survey on Qualtrics or a similar service.

Revised 07/12/2022

using a workstation in the lab area. This step should take no more than 15 minutes, giving a total of 60 minutes for the participant. In between sessions, the VR equipment will be cleaned and sanitized.

13C. List all data collection instruments used in this project, in the order they appear in Appendix C.

(e.g., surveys and questionnaires in the format that will be presented to participants, educational tests, data collection sheets, interview questions, audio/video taping methods etc.)

Items in Appendix C: (1) The pre-survey, a bespoke survey for this project. (2) The post-survey for participants using the traditional desktop environment, which includes: (2a) The NASA Task Load Index (TLX) [7], raw, to gauge workload; and (2b) Brooke's System Usability Scale [8], to gauge user satisfaction. (3) The post-survey for participants using the basic virtual reality environment, which includes: (3a) The NASA Task Load Index (TLX) [7], to gauge cognitive load; (3b) Brooke's System Usability Scale [8], to gauge user satisfaction; and (3c) applicable elements from Witmer, Jerome, and Sanger's updated Presence Questionnaire [9], to gauge immersiveness. (4) The post-survey for participants using the enhanced virtual reality environment, which includes: (4a) The NASA Task Load Index (TLX) [7], to gauge cognitive load; (4b) Brooke's System Usability Scale [8], to gauge user satisfaction; and (4c) applicable elements from Witmer, Jerome, and Sanger's updated Presence Questionnaire [9], to gauge immersiveness. (4d) a series of Likert questions about the value of features in the environment. All surveys will be conducted using a suitable web-based service such as Qualtrics.

13D. Data analysis: Describe how data will be analyzed. If a data collection form (DCF) will be used, submit a copy of the DCF.

Workload, System Usability, and Presence will be compared between environments using a statistical hypothesis test. Accuracy and speed of completing the reverse engineering challenge will be recorded in a spreadsheet with four fields: Coded participant ID, Environment, Accuracy, Speed, and then will be compared between environments using a statistical hypothesis test. Logged data will be analyzed for trends and patterns such as most frequently-used commands or features, or hotspot activity locations in the virtual space.

13E. List any drugs, medications, supplements, or imaging agents that participants will ingest/ receive during participation in the study or indicate not applicable (N/A).

N/A

14. Risks & Discomforts: List and describe all the risks participants may encounter in this research including risks from item 6d of this form, in this research. If deception will be part of the study, provide the rationale for the deception, describe the debriefing process, and attach a copy of the debriefing form that will be used as Appendix D. (Examples of possible risks are in section #6C)

Potential risks include: physical discomfort (such as simulator sickness or headaches from the VR equipment), spread of germs (viruses and bacteria) via the equipment, and confidentiality breach.

15. Precautions / Minimization of Risks

15A. Identify and describe all precautions that will be taken to eliminate or reduce risks listed in items 6.c. and 14. If participants can be classified as a "vulnerable" population, describe additional safeguards that will be used to assure the ethical treatment of vulnerable individuals. If applicable, submit a copy of any emergency plans/procedures and medical referral lists in Appendix D. (Sample documents can be found online at <https://cws.auburn.edu/OVPR/pm/compliance/irb/sampledocs> precautions)

Participants will either be seated at a desktop workstation computer or using virtual reality (VR) equipment. Those using VR equipment will stand and wear a headset and hold a controller in each hand in a contained area, free of obstacles, monitored by the researchers. Before participation, VR users will be screened using the questions listed in

Revised 07/12/2022

Appendix D. Participants will be allowed to take breaks, including temporarily removing the VR equipment, during the task. Participants will be told to immediately inform the researcher if they feel dizzy or uncomfortable with the VR equipment. If discomfort is expressed, the activity will cease immediately for that participant. They will be monitored by a researcher to ensure their well-being. If discomfort persists, appropriate recommendations for further assistance will be provided. In between sessions, the VR equipment will be cleaned and sanitized. To protect confidentiality, data from surveys will be coded, and participants' Auburn usernames will not be included. The results will only feature their survey responses and task performance alongside a code representing their identity.

15B. If the internet, mobile apps, or other electronic means will be used to collect data, describe confidentiality and/or security precautions that will be used to protect (or not collect) identifiable data? Include protections used during collection of data, transfer of data, and storage of data. If participant data may be obtained and/or stored by apps during the study, describe.

Survey data, including identifiable data, will be collected with Qualtrics or a similar tool. The tool will be one approved by Auburn for which the researchers have a valid license to use. All data sent to or received by this tool will be over encrypted communications. Each participant will be assigned a coded ID for the purposes of this study. Any data downloaded from the survey tool will only use the coded ID. As the participant works to solve a reverse engineering challenge in the desktop environment, actions such as executing command line actions will be logged associated with that ID. As the participant works to solve a reverse engineering challenge in the virtual reality environment, the participant's telemetry (position and orientation of head and hands) and actions will be logged associated with that ID. Participant performance on the challenge will be stored in a file associated with the coded ID. All data in stored as files will be stored in a secure folder, identified by the coded ID, on an Auburn-licensed secure service (BOX), with access limited to those authorized to review the data.

15C. Does this research include purchase(s) that involve technology hardware, software or online services?

☒ YES ☐ NO

If YES:

- A. Provide the name of the product VIVE Pro 2
and the manufacturer of the product HTC
- B. Briefly describe use of the product in the proposed human subject's research.
The HTC VIVE Pro 2 provides the interface to the immersive virtual reality environments employed in this study.
- C. To ensure compliance with AU's Electronic and Information Technology Accessibility Policy, contact AU IT Vendor Vetting team at vetting@auburn.edu to learn the vendor registration process (prior to completing the purchase).
- D. Include a copy of the documentation of the approval from AU Vetting with the revised submission.
Attached as the final pages in this document.

15D. Additional Safeguards

Will DEXA, pQCT, or other devices which emit radiation be used? ☐ Yes ☒ No

If yes, the IRB will notify the Auburn Department of Risk Management and Safety, who will contact the Alabama Department of Public Health (ADPH) and secure approval. Research which includes device(s) which emit radiation may NOT be initiated NOR will IRB stamped consent documents be issued until the IRB is notified of ADPH approval.

Will a Certificate of Confidentiality (CoC) issued by NIH be obtained ☐ Yes ☒ No If yes, include CoC language in consent documents and include the documentation of CoC approval. Research which includes a CoC may not be initiated NOR will IRB stamped consent documents be issued until the IRB is notified of CoC approval. [AU Required CoC Language](#)

Is the study a [clinical trial](#)? ☐ Yes ☒ No

If yes, provide the National Clinical Trial (NCT) # [Click or tap here to enter text.](#) and include required clinical trial information in all consent documents. [AU Clinical Trial Information](#)

Revised 07/12/2022

16. Benefits

16A. List all realistic direct benefits participants can expect by participating in this study. (Compensation is not a benefit) If participants will not directly benefit check here. ☐

Participants will gain experience in solving a binary reverse engineering problem, and additional experience using a novel virtual reality environment. Since binary code is the foundation of digital computing and much of computer science, this experience will supplement topics relevant to the target participant pool of Auburn engineering students.

16B. List realistic benefits for the general population that may be generated from this study.

Many analytical problems in general, and binary reverse engineering in particular, tax human cognition, yet still require a human to solve. This study measures the effectiveness of one approach to augmenting human cognition using immersive virtual reality. The study is narrowly-focused on the specific problem of reverse engineering, but the findings may generalize to other problem areas. We will share our findings with the scientific community for their consideration to apply them in their own problem areas with the ultimate goal of making hard analytical problems easier and/or faster to solve.

17. Protection of Data

17A. Data are collected:

- ☐ Anonymously with no direct or indirect coding, link, or awareness by key personnel of who participated in the study (skip to item E)
- ☐ Confidentially, but without a link to participant's data to any identifying information (collected as "confidential" but recorded and analyzed "anonymous") (Skip to item E).
- ☒ Confidentially with collection and protection of linkages to identifiable information.

17B. If data are collected with identifiers and coded or as coded or linked to identifying information, describe the identifiers and how identifiers are linked to participants' data.

The participant's Auburn username will be collected in the pre-survey and post-session survey in order to link them. The responses will be assigned a coded ID in place of the username before any data is moved to the secure shared folder.

17C. Provide the rationale for need to code participants' data or link the data with identifying information.

Due to the nature of the skills required to complete this activity, it is expected that participants of widely-varying experience in binary reverse engineering will sign up for this study. Participants self-assess their experience levels in the pre-survey, and in order to understand the relationship between experience with performance and its impacts on our results, we need to link the pre-survey data with the performance and post-session survey data.

17D. Describe how and where identifying data and/or code lists will be stored. (Building, room number, AU BOX?) **Describe how the location where data is stored will be secured. For electronic data, describe security measures. If applicable, describe where IRB-approved and participant signed consent documents will be kept on campus for 3 years after the study ends.**

Electronic data will be collected through surveys using an approved Auburn-licensed service such as Qualtrics. Any data downloaded from the service will use a coded participant ID and will be stored in a secure folder on an Auburn-licensed service (BOX) with access limited to those authorized to review that data as identified in item 17F. Personal identifications will remain with the online survey tool and not be downloaded to the secure shared folder. Paper consent forms will be stored in a lockable container in Dr. Mulder's office for 3 years.

Revised 07/12/2022

17E. Describe how and where data will be stored (e.g., hard copy, audio/ visual files, electronic data, etc.), and how the location where data is stored is separated from identifying data and will be secured. For electronic data, describe security. Note use of a flash drive or portable hard drive is not appropriate if identifiable data will be stored; rather, identifying participant data must be stored on secured servers.

Identifying data will be collected through surveys using an approved Auburn-licensed service such as Qualtrics. Any data downloaded from the survey service will use a coded participant ID and not contain personal identifiers. Any other data collected will be associated with the coded ID only. All collected data required for the analysis, which will only use coded IDs, will be stored in a secure folder on an Auburn-licensed service (BOX) with access limited to those authorized to review that data as identified in item 17F. No personal identifiers will be stored in the shared folder.

17F. List the names of all who will have access to participants' data? (If a student PI, the faculty advisor must have full access and be able to produce study data in the case of a federal or institutional audit.)

Dennis Brown (student PI), Dr. Samuel Mulder (faculty advisor), Kevan Baker, Julian Bauer

17G. When is the latest date that identifying information or links will be retained and how will that information or links be destroyed? (Check here if only anonymous data will be retained ☐)

Identifying data will be stored for 3 years. Non-identifying data will be stored indefinitely.

Version Date: 7/22/2024

Revised 07/12/2022

Consent Forms

There are two forms: one contained within the pre-survey, and one presented at the time of the participant's in-person session.

These forms are on the following pages (inserted into the final PDF).

PRE-SURVEY CONSENT FORM



NOTE: DO NOT SIGN THIS DOCUMENT UNLESS AN IRB APPROVAL STAMP WITH CURRENT DATES HAS BEEN APPLIED TO THIS DOCUMENT.

**INFORMED CONSENT
for the research study
“Impacts of Immersive Virtual Reality on the Binary Reverse Engineering Process”**

Thank you for your interest in participating in this study, which aims to measure the effectiveness of different configurations of tools designed to facilitate the process of reverse engineering a binary program. These configurations will include a traditional computer desktop environment and two immersive virtual reality (VR) environments.

This study is being conducted by PhD student Dennis Brown under the direction of Dr. Samuel Mulder in the Program Understanding Lab, Computer Science and Software Engineering department, Auburn University.

Participants should be Auburn University students age 19 or older. Participants ineligible to use the virtual reality environment include those with cold, flu, headache, migraines, earache, severe hand deformity or pain, pregnancy, pre-existing binocular vision abnormalities (e.g., strabismus (crossed eye) and amblyopia (lazy eye)), or other serious medical conditions. Ineligible persons should discontinue this process at this point.

What will be involved if you participate? In this survey, you will be asked for personal information including your Auburn username and a self-rating of your experience in relevant topics. On the form to sign up for a time slot, you will be asked for your Auburn email. This consent form covers your participation in only this survey and time slot sign up. At the start of the session for which you’ve signed up, you will be asked to provide a second consent to continuing your participation. In that session you will attempt to solve a reverse engineering challenge in one of the three environments (the environment to be used is randomly assigned to the participant), and then complete a post-survey about your experience. The entire experiment includes completing this pre-survey, signing up for a time slot, appearing at the study location at your appointed time, attempting the reverse engineering tasks, and completing the post-survey.

Participant’s Initials: _____

The Auburn University Institutional
Review Board has approved this
Document for use from
07/23/2024 to -----
Protocol # 24-668 EP 2407

Page 1

Participants who complete the entire experiment will receive a \$20 e-gift card sent to the Auburn email address you provide on the time slot sign up form.

Are there any risks or discomforts? This consent form is specifically for this survey and time slot sign up, in which you are asked personal information as listed above. The primary risk is confidentiality. All electronic data you provide or that is collected about your participation is kept confidential and stored in a secure folder accessible only to those authorized to view it.

Are there any benefits to yourself or others? Your benefit will be an opportunity to learn more about reverse engineering. The benefit to the researchers is to provide empirical data that helps us prove or disprove hypotheses about using immersive virtual reality for reverse engineering, the results of which will be shared with the scientific community and inform future research directions.

Will you receive compensation for participating? Participants who complete the experiment, as defined above, will receive a \$20 e-gift card from a major retailer. Note that completion includes attempting to solve the challenge but does not require a full and correct solution.

Are there any costs? Participants will have no monetary cost for participating in this study.

If you change your mind about participating, you can withdraw at any time during the study. Your participation is voluntary and will not affect your future relations with the study investigators or Auburn University. If you decide to withdraw, your data will be promptly deleted from the database, and at this stage, you can do so by simply closing your browser.

Any information obtained in connection with this study will remain confidential. Data will be securely stored on an Auburn University-approved server with password-protected access limited to researchers involved in the study. Information gathered may be published in professional journals or presented at meetings, but will not personally identify participants. By signing below, you agree to allow your personal information from the pre-survey to be used for matching responses and for scheduling your experiment. Only IRB-approved individuals will have access to this information, and it will not be downloaded or shared otherwise.

If you have questions about the study, please contact Dennis Brown at dgb0028@auburn.edu or Dr. Samuel Mulder at szm0211@auburn.edu.

If you have questions about your rights as a research participant, you may contact the Auburn University Office of Human Subjects Research or the Institutional Review Board by phone (334) 844-5966 or e-mail at IRBadmin@auburn.edu or IRBChair@auburn.edu.

Participant's Initials: _____

The Auburn University Institutional
Review Board has approved this
Document for use from
07/23/2024 to -----
Protocol # 24-668 EP 2407

Page 2

**HAVING READ THE INFORMATION PROVIDED, YOU MUST DECIDE WHETHER OR NOT
YOU WISH TO PARTICIPATE IN THIS RESEARCH STUDY. YOUR SIGNATURE INDICATES
YOUR WILLINGNESS TO PARTICIPATE.**

Participant's signature Date

Participant's printed name

Signature of investigator obtaining consent Date

Investigator's printed name

Version date: June 24, 2024

The Auburn University Institutional
Review Board has approved this
Document for use from
07/23/2024 to -----
Protocol # 24-668 EP 2407

Participant's Initials: _____

SESSION CONSENT FORM



NOTE: DO NOT SIGN THIS DOCUMENT UNLESS AN IRB APPROVAL STAMP WITH CURRENT DATES HAS BEEN APPLIED TO THIS DOCUMENT.

**INFORMED CONSENT
for the research study
“Impacts of Immersive Virtual Reality on the Binary Reverse Engineering Process”**

Thank you for your interest in participating in this study, which aims to measure the effectiveness of different configurations of tools designed to facilitate the process of reverse engineering a binary program. These configurations will include a traditional computer desktop environment and two immersive virtual reality (VR) environments.

This study is being conducted by PhD student Dennis Brown under the direction of Dr. Samuel Mulder in the Program Understanding Lab, Computer Science and Software Engineering department, Auburn University.

Participants should be Auburn University students age 19 or older. Participants ineligible to use the virtual reality environment include those with cold, flu, headache, migraines, earache, severe hand deformity or pain, pregnancy, pre-existing binocular vision abnormalities (e.g., strabismus (crossed eye) and amblyopia (lazy eye)), or other serious medical conditions.

What will be involved if you participate? You will be randomly assigned to one of three possible reverse engineering tool environments, a traditional computer desktop environment, a basic virtual reality (VR) environment, and an enhanced VR environment. You will attempt one reverse engineering challenge in your assigned environment. You will be allotted up to 30 minutes to solve the challenge. You can take breaks and end the attempt at any time. In the traditional environment, participants will use reverse engineering tools in a common desktop windowed environment controlled by keyboard and mouse. In the VR environments, they will wear an HTC Vive Pro 2 VR headset and use two HTC VR controllers. During your attempt, our system will log data about your actions. After that attempt, you will complete an online survey about system usability, workload, immersion (if using a VR environment), and feature rating (if using the enhanced VR environment). Complete sessions, from consent to performing the task to answering a survey, should take no longer than 60 minutes per participant.

Participant's Initials: _____

The Auburn University Institutional
Review Board has approved this
Document for use from
07/23/2024 to _____
Protocol # 24-668 EP 2407

Page 1

Are there any risks or discomforts? One primary risk associated with your participation is physical discomfort, such as motion/simulator sickness, headache, or general discomfort wearing and using the virtual reality equipment. Another risk is confidentiality. All electronic data you provide or that is collected about your participation is kept confidential and stored in a secure folder accessible only to those authorized to view it.

Are there any benefits to yourself or others? Your benefit will be an opportunity to learn more about reverse engineering. The benefit to the researchers is to provide empirical data that helps us prove or disprove hypotheses about using immersive virtual reality for reverse engineering, the results of which will be shared with the scientific community and inform future research directions.

Will you receive compensation for participating? Participants who complete the experiment, as defined above, will receive a \$20 e-gift card from a major retailer. Note that completion includes attempting to solve the challenge but does not require a full and correct solution.

Are there any costs? Participants will have no monetary cost for participating in this study.

If you change your mind about participating, you can withdraw at any time during the study. Your participation is voluntary and will not affect your future relations with the study investigators or Auburn University. If you decide to withdraw, your data will be promptly deleted from the database, and at this stage, you can do so by simply closing your browser.

Any information obtained in connection with this study will remain confidential. Data will be securely stored on an Auburn University-approved server with password-protected access limited to researchers involved in the study. Information gathered may be published in professional journals or presented at meetings, but will not personally identify participants. By signing below, you agree to allow your personal information from the pre-survey to be used for matching responses and for scheduling your experiment. Only IRB-approved individuals will have access to this information, and it will not be downloaded or shared otherwise.

If you have questions about the study, please contact Dennis Brown at dgb0028@auburn.edu or Dr. Samuel Mulder at szm0211@auburn.edu.

If you have questions about your rights as a research participant, you may contact the Auburn University Office of Human Subjects Research or the Institutional Review Board by phone (334) 844-5966 or e-mail at IRBAdmin@auburn.edu or IRBChair@auburn.edu.

Participant's Initials: _____

The Auburn University Institutional
Review Board has approved this
Document for use from
07/23/2024 to -----
Protocol # 24-668 EP 2407

Page 2

HAVING READ THE INFORMATION PROVIDED, YOU MUST DECIDE WHETHER OR NOT YOU WISH TO PARTICIPATE IN THIS RESEARCH STUDY. YOUR SIGNATURE INDICATES YOUR WILLINGNESS TO PARTICIPATE.

Participant's signature

Date

Participant's printed name

Signature of investigator obtaining consent

Date

Investigator's printed name

Version date: June 24, 2024

The Auburn University Institutional
Review Board has approved this
Document for use from
07/23/2024 to -----
Protocol # 24-668 EP 2407

Participant's Initials: _____

Revised 07/12/2022

Appendix A (Reference List)

- [1] H. G. Rice, "Classes of recursively enumerable sets and their decision problems," *Journal of Symbolic Logic*, vol. 19, no. 2, pp. 121–122, 1954.
- [2] D. Crockford, *JavaScript: The Good Parts*. O'Reilly Media, Inc., 2008.
- [3] A. Elliott, B. Peiris, and C. Parnin, "Virtual reality in software engineering: Affordances, applications, and challenges," in *2015 IEEE/ACM 37th IEEE International Conference on Software Engineering*, vol. 2, 2015, pp. 547–550.
- [4] J. Dominic, B. Tubre, J. Houser, C. Ritter, D. Kunkel, and P. Rodeghero, *Program Comprehension in Virtual Reality*. New York, NY, USA: Association for Computing Machinery, 2020, p. 391–395. [Online]. Available: <https://doi.org/10.1145/3387904.3389287>
- [5] S. Romano, N. Capece, U. Erra, G. Scanniello, and M. Lanza, "On the use of virtual reality in software visualization: The case of the city metaphor," *Information and Software Technology*, vol. 114, pp. 92–106, 2019. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0950584919301405>
- [6] A. Hoff, L. Gerling, and C. Seidl, "Utilizing software architecture recovery to explore large-scale software systems in virtual reality," in *2022 Working Conference on Software Visualization (VISSOFT)*, 2022, pp. 119–130.
- [7] S. G. Hart, "Nasa task load index (tlx). Volume 1.0; paper and pencil package," 1986. [Online]. Available: <https://ntrs.nasa.gov/citations/20000021488>
- [8] J. Brooke, *SUS – a quick and dirty usability scale*. Boca Raton, FL, USA: CRC Press, 01 1996, pp. 189–194.
- [9] B. G. Witmer, C. J. Jerome, and M. J. Singer, "The Factor Structure of the Presence Questionnaire," *Presence: Teleoperators and Virtual Environments*, vol. 14, no. 3, pp. 298–312, 06 2005. [Online]. Available: <https://doi.org/10.1162/105474605323384654>
- [10] M. Verkuyl, K. Taplay, L. Attack, M. Boulet, N. Dubois, S. Goldsworthy, T. Merwin, T. Willett, and T. Job, *Virtual Simulation: An Educator's Toolkit*. eCampusOntario Open Library. [Online]. Available: <https://ecampusontario.pressbooks.pub/vlsvstoolkit/>

Revised 07/12/2022

Appendix B (Recruitment)

Recruitment email text:

Auburn engineering students,

Do you love low-level coding, solving puzzles, and helping people?

My name is Dennis Brown and I am a PhD student in the Computer Science and Software Engineering department, working with my advisor, Dr. Samuel Mulder, in the Program Understanding Lab.

We are conducting a study “Impacts of Immersive Virtual Reality on the Binary Reverse Engineering Process” and we seek volunteers to help us by participating in the study.

This study aims to measure the effectiveness of different configurations of tools designed to facilitate the process of reverse engineering a binary program. These configurations will include a traditional computer desktop environment and two immersive virtual reality environments.

In this study, volunteers will complete a very short pre-survey and then sign up for a 1-hour time slot. During your scheduled time, you will be asked to attempt to solve a reverse engineering challenge in one of the three environments (the environment to be used is randomly assigned to the participant) and complete a post-survey about your experience.

Participants who complete all steps of the experiment will receive a \$20 e-gift card from a major retailer, where completion includes attempting to solve the challenge but does not require a full and correct solution.

For any questions about this study, please contact Dennis Brown at dgb0028@auburn.edu or Dr. Samuel Mulder at szm0211@auburn.edu.

To volunteer, please visit this URL and follow the instructions: <URL TBD>

Thank you,

Dennis Brown

The Auburn University Institutional
Review Board has approved this
Document for use from
07/23/2024 to -----
Protocol # 24-668 EP 2407

Revised 07/12/2022

Appendix C (Data Collection)

There are four surveys:

1. Pre-survey
2. Post-survey for the traditional desktop environment
3. Post-survey for the basic virtual reality environment
4. Post-survey for the enhanced virtual reality environment

These forms are on the following pages (inserted into the final PDF).

The Auburn University Institutional
Review Board has approved this
Document for use from
07/23/2024 to -----
Protocol # 24-668 EP 2407

INFORMATION LETTER

Thank you for your interest in participating in this study, which aims to measure the effectiveness of different configurations of tools designed to facilitate the process of reverse engineering a binary program. These configurations will include a traditional computer desktop environment and two immersive virtual reality (VR) environments.

This study is being conducted by PhD student Dennis Brown under the direction of Dr. Samuel Mulder in the Program Understanding Lab, Computer Science and Software Engineering department, Auburn University.

Participants should be Auburn University students age 19 or older. Participants ineligible to use the virtual reality environment include those with cold, flu, headache, migraines, earache, severe hand deformity or pain, pregnancy, pre-existing binocular vision abnormalities (e.g., strabismus (crossed eye) and amblyopia (lazy eye)), or other serious medical conditions. Ineligible persons should discontinue this process at this point.

What will be involved if you participate? In this survey, you will be asked for personal information including your Auburn username and a self-rating of your experience in relevant topics. On the form to sign up for a time slot, you will be asked for your Auburn email. This consent form covers your participation in only this survey and time slot sign up. At the start of the session for which you've signed up, you will be asked to provide a second consent to continuing your participation. In that session you will attempt to solve a reverse engineering challenge in one of the three environments (the environment to be used is randomly assigned to the participant), and then complete a post-survey about your experience. The entire experiment includes completing this pre-survey, signing up for a time slot, appearing at the study location at your appointed time, attempting the reverse engineering tasks, and completing the post-survey. Participants who complete the entire experiment will receive a

\$20 e-gift card sent to the Auburn email address you provide on the time slot sign up form.

Are there any risks or discomforts? This consent form is specifically for this survey and time slot sign up, in which you are asked personal information as listed above. The primary risk is confidentiality. All electronic data you provide or that is collected about your participation is kept confidential and stored in a secure folder accessible only to those authorized to view it.

Are there any benefits to yourself or others? Your benefit will be an opportunity to learn more about reverse engineering. The benefit to the researchers is to provide empirical data that helps us prove or disprove hypotheses about using immersive virtual reality for reverse engineering, the results of which will be shared with the scientific community and inform future research directions.

Will you receive compensation for participating? Participants who complete the experiment, as defined above, will receive a \$20 e-gift card from a major retailer. Note that completion includes attempting to solve the challenge but does not require a full and correct solution.

Are there any costs? Participants will have no monetary cost for participating in this study.

If you change your mind about participating, you can withdraw at any time during the study. Your participation is voluntary and will not affect your future relations with the study investigators or Auburn University. If you decide to withdraw, your data will be promptly deleted from the database, and at this stage, you can do so by simply closing your browser.

Any information obtained in connection with this study will remain confidential. Data will be securely stored on an Auburn University-approved server with password-protected access limited to researchers involved in the study. Information gathered may be published in professional journals or presented at meetings, but will not personally identify participants. By signing below, you agree to allow your personal information from the pre-survey to be used for matching responses and for scheduling your experiment. Only IRB-approved individuals will have access to this information, and it will not be downloaded or shared otherwise.

If you have questions about the study, please contact Dennis Brown at dgb0028@auburn.edu or Dr. Samuel Mulder at szm0211@auburn.edu.

The Auburn University Institutional Review Board has approved this Document for use from 07/23/2024 to ----- Protocol # 24-668 EP 2407
--

If you have questions about your rights as a research participant, you may contact the Auburn University Office of Human Subjects Research or the Institutional Review Board by phone (334) 844-5966 or e-mail at IRBadmin@auburn.edu or IRBChair@auburn.edu.

HAVING READ THE INFORMATION PROVIDED, YOU MUST DECIDE WHETHER OR NOT YOU WISH TO PARTICIPATE IN THIS RESEARCH STUDY. COMPLETING AND SUBMITTING THIS FORM INDICATES YOUR WILLINGNESS TO PARTICIPATE.

Click this link to download a copy of this consent form for your records: <URL TBD>

Please enter your Auburn username, e.g., abc0003

Please enter your major field of study

Please rate yourself in each of the following items:

My level of expertise regarding software development is:

None

Moderate

High

☐☐☐☐☐

My level of expertise regarding computer architecture is:

None

Moderate

High

☐☐☐☐☐

The Auburn University Institutional
Review Board has approved this
Document for use from
07/23/2024 to -----
Protocol # 24-668 EP 2407

My level of expertise regarding how compilers are constructed is:

None		Moderate		High
☐	☐	☐	☐	☐

My level of expertise regarding reverse engineering of binary code is:

None		Moderate		High
☐	☐	☐	☐	☐

My level of experience as a *user* of virtual reality is:

None		Moderate		High
☐	☐	☐	☐	☐

My level of expertise as a *designer or developer* of virtual reality is:

None		Moderate		High
☐	☐	☐	☐	☐

→

Tools ▾

Saved at 10:22 AM

Draft



Preview

Publish

immersiveRE-desktop-mockup



ExpertReview score

Fair

▼ ID

Q1

Please enter your Auburn username, e.g., abc0003

Import from library

Add new question

Add Block

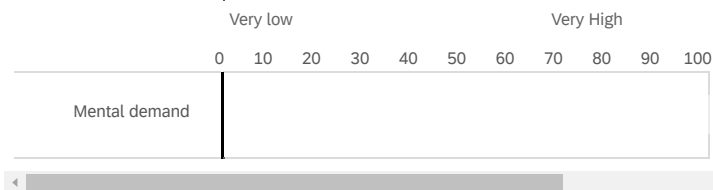
▼ Task Load Index

Q2

Task Load: Please read the following questions and rate yourself in each category by sliding the control between 0 and 100.

Q3

Mental demand: How much mental and perceptual activity was required? Was the task easy or demanding, simple or complex?

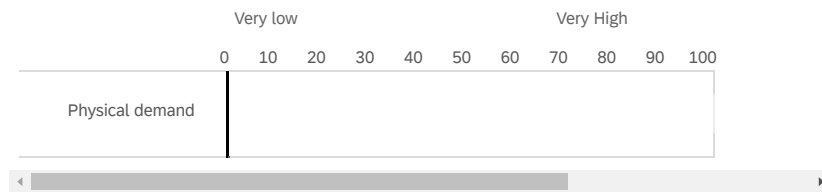


The Auburn University Institutional
Review Board has approved this
Document for use from

07/23/2024 to -----
Protocol # 24-668 EP 2407

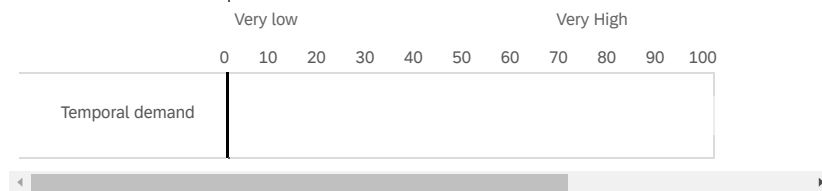
Q4

Physical demand: How much physical activity was required? Was the task easy or demanding, slack or strenuous?



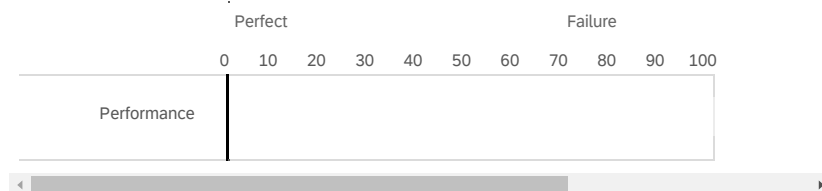
Q5

Temporal demand: How much time pressure did you feel due to the pace at which the tasks or task elements occurred? Was the pace slow or rapid?



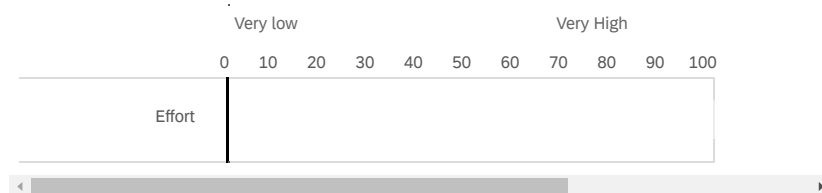
Q6

Performance: How successful were you in performing the task? How satisfied were you with your performance?



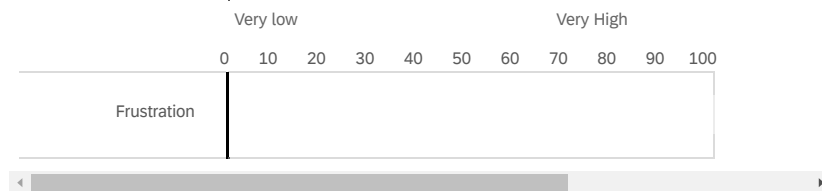
Q7

Effort: How hard did you have to work (mentally and physically) to accomplish your level of performance?



Q8

Frustration: How irritated, stressed, and annoyed versus content, relaxed, and complacent did you feel during the task?



Page Break

 Import from library

Add new question

Add Block

System Usability Survey

Q9

System Usability: Please read each statement and select your level of agreement.

Q10



I think that I would like to use this system frequently

Strongly disagree

Strongly agree



Q11



I found the system unnecessarily complex

Strongly disagree

Strongly agree



Q12



I thought the system was easy to use

Strongly disagree

☐☐☐☐

Strongly agree

☐

Q13



I think that I would need the support of a technical person to be able to use this system

Strongly disagree

☐☐☐☐

Strongly agree

☐

Q14



I found the various functions in this system were well integrated

Strongly disagree

☐☐☐☐

Strongly agree

☐

Q15



I thought there was too much inconsistency in this system

Strongly disagree

☐☐☐☐

Strongly agree

☐

Q16



I would imagine that most people would learn to use this system very quickly

Strongly disagree

☐☐☐☐

Strongly agree

☐

Q17



I found the system very cumbersome to use

Strongly disagree

☐☐☐☐

Strongly agree

☐

Q18



I felt very confident using the system

Strongly disagree

☐☐☐☐

Strongly agree

☐

Q19



I needed to learn a lot of things before I could get going with this system

Strongly disagree

☐☐☐☐

Strongly agree

☐

 Import from library

Add new question

[Add Block](#)

End of Survey

We thank you for your time spent taking this survey.

Your response has been recorded.



immersiveRE-VR1-mockup

ExpertReview score

Fair

▼ ID

Q1

...

Please enter your Auburn username, e.g., abc0003

▲

Import from library

Add new question

Add Block

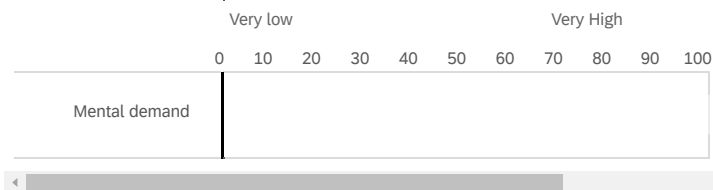
▼ Task Load Index

Q2

Task Load: Please read the following questions and rate yourself in each category by sliding the control between 0 and 100.

Q3

Mental demand: How much mental and perceptual activity was required? Was the task easy or demanding, simple or complex?



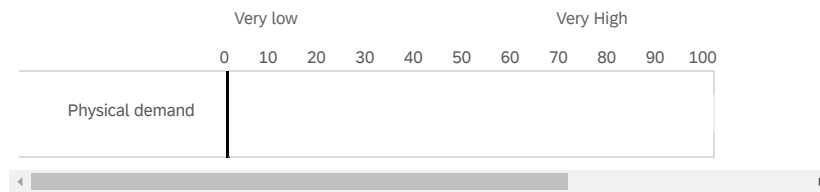
The Auburn University Institutional
Review Board has approved this
Document for use from

07/23/2024 to -----

Protocol # 24-668 EP 2407

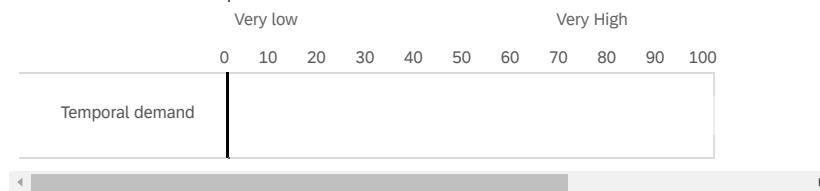
Q4

Physical demand: How much physical activity was required? Was the task easy or demanding, slack or strenuous?



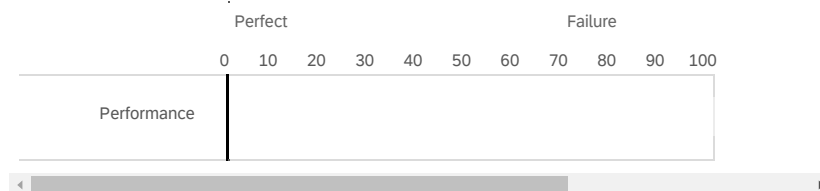
Q5

Temporal demand: How much time pressure did you feel due to the pace at which the tasks or task elements occurred? Was the pace slow or rapid?



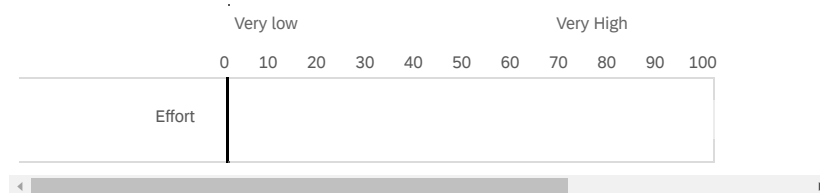
Q6

Performance: How successful were you in performing the task? How satisfied were you with your performance?



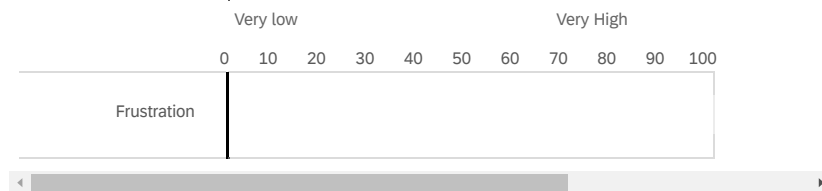
Q7

Effort: How hard did you have to work (mentally and physically) to accomplish your level of performance?



Q8

Frustration: How irritated, stressed, and annoyed versus content, relaxed, and complacent did you feel during the task?



Page Break

 Import from library

Add new question

Add Block

▼ System Usability Survey

Q9

System Usability: Please read each statement and select your level of agreement.

Q10



I think that I would like to use this system frequently

Strongly disagree

Strongly agree

☐☐☐☐☐

Q11



I found the system unnecessarily complex

Strongly disagree

Strongly agree

☐☐☐☐☐

Q12



I thought the system was easy to use

Strongly disagree

☐☐☐☐

Strongly agree

☐

Q13



I think that I would need the support of a technical person to be able to use this system

Strongly disagree

☐☐☐☐

Strongly agree

☐

Q14



I found the various functions in this system were well integrated

Strongly disagree

☐☐☐☐

Strongly agree

☐

Q15



I thought there was too much inconsistency in this system

Strongly disagree

☐☐☐☐

Strongly agree

☐

Q16



I would imagine that most people would learn to use this system very quickly

Strongly disagree

☐☐☐☐

Strongly agree

☐

Q17



I found the system very cumbersome to use

Strongly disagree

☐☐☐☐

Strongly agree

☐

Q18



I felt very confident using the system

Strongly disagree

Strongly agree

☐☐☐☐☐

Q19



I needed to learn a lot of things before I could get going with this system

Strongly disagree

Strongly agree

☐☐☐☐☐

 Import from library

Add new question

[Add Block](#)



Presence

Q20

Presence: Please read each question and choose the appropriate level.

Q21



How much were you able to control events?

Not at all

Moderately

Completely

☐☐☐☐☐☐☐

Q22



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

Not at all

Moderately

Completely

☐☐☐☐☐☐☐

Q23



How natural did your interactions with the environment seem?

Not at all

☐☐☐

Moderately

☐☐☐

Completely

☐

Q24



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

Not at all

☐☐☐

Moderately

☐☐☐

Completely

☐

Q25



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

Not at all

☐☐☐

Moderately

☐☐☐

Completely

☐

Q26



How compelling was your sense of objects moving through space?

Not at all

☐☐☐

Moderately

☐☐☐

Completely

☐

Q27



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

Not at all

☐☐☐

Moderately

☐☐☐

Completely

☐

Q28



Were you able to anticipate what would happen next in response to the actions that you performed?

Not at all

☐☐☐

Moderately

☐☐☐

Completely

☐

Q29



How completely were you able to actively survey or search the environment using vision?

Not at all

☐☐☐

Moderately

☐☐☐

Completely

☐

Q30



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

Not at all

☐☐☐

Moderately

☐☐☐

Completely

☐

Q31



How well could you move or manipulate objects in the virtual environment?

Not at all

☐☐☐

Moderately

☐☐☐

Completely

☐

Q32



How involved were you in the virtual environment experience?

Not at all

☐☐☐

Moderately

☐☐☐

Completely

☐

Q33



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

Not at all

☐☐☐

Moderately

☐☐☐

Completely

☐

Q34



How quickly did you adjust to the virtual environment experience?

Not at all

☐☐☐

Moderately

☐☐☐

Completely

☐

Q35



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

Not at all

☐☐☐

Moderately

☐☐☐

Completely

☐

Q36



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

Not at all

☐☐☐

Moderately

☐☐☐

Completely

☐

Q37



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

Not at all

☐☐☐

Moderately

☐☐☐

Completely

☐

Q38



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

Not at all

☐☐☐

Moderately

☐☐☐

Completely

☐

Q39



How completely were your senses engaged in this experience?

Not at all

☐☐☐

Moderately

☐☐☐

Completely

☐

Q40



How easy was it to identify objects through physical interaction, like touching an object, walking over a surface, or bumping into a wall or object?

Not at all

☐☐☐

Moderately

☐☐☐

Completely

☐

Q41



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

Not at all

☐☐☐

Moderately

☐☐☐

Completely

☐

Q42



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

Not at all

☐☐☐

Moderately

☐☐☐

Completely

☐

Q43



Was the information provided through different senses in the virtual environment (e.g., vision, hearing, touch) consistent?

Not at all

☐☐☐

Moderately

☐☐☐

Completely

☐

Import from library

Add new question

[Add Block](#)

End of Survey

We thank you for your time spent taking this survey.

Your response has been recorded.



immersiveRE-VR2-mockup

ExpertReview score

Fair

▼ ID

Q1

...

Please enter your Auburn username, e.g., abc0003

▲

Import from library

Add new question

[Add Block](#)

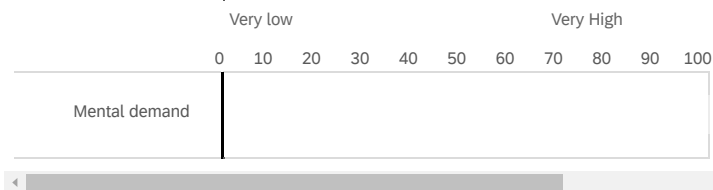
▼ Task Load Index

Q2

Task Load: Please read the following questions and rate yourself in each category by sliding the control between 0 and 100.

Q3

Mental demand: How much mental and perceptual activity was required? Was the task easy or demanding, simple or complex?



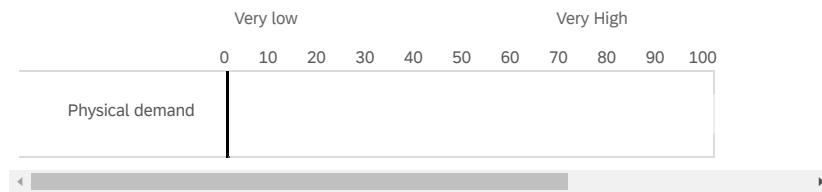
The Auburn University Institutional
Review Board has approved this
Document for use from

07/23/2024 to -----

Protocol # 24-668 EP 2407

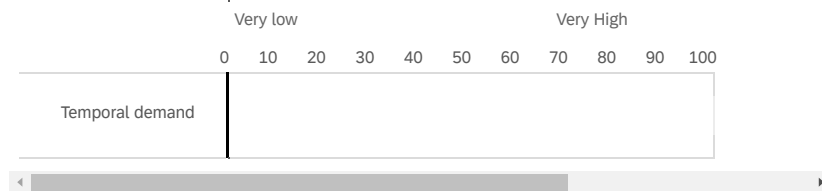
Q4

Physical demand: How much physical activity was required? Was the task easy or demanding, slack or strenuous?



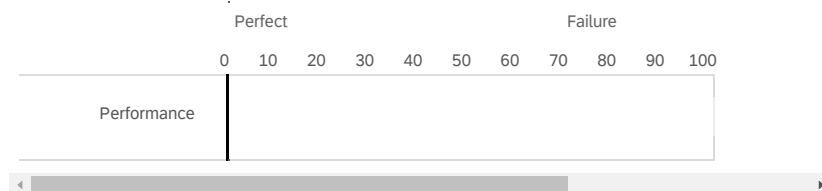
Move 1 block

Temporal demand: How much time pressure did you feel due to the pace at which the tasks or task elements occurred? Was the pace slow or rapid?



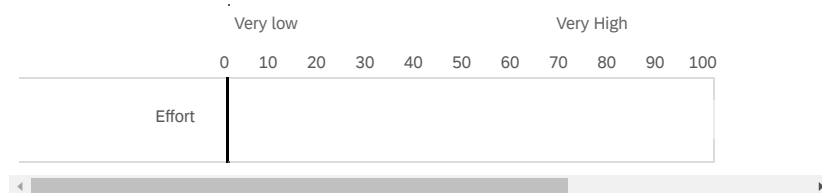
Q6

Performance: How successful were you in performing the task? How satisfied were you with your performance?



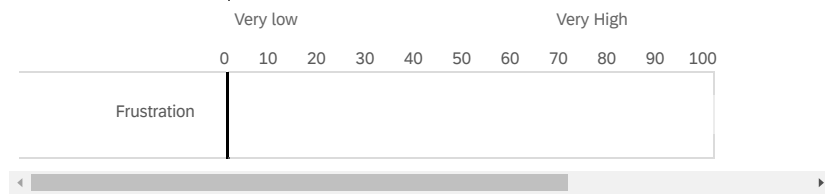
Q7

Effort: How hard did you have to work (mentally and physically) to accomplish your level of performance?



Q8

Frustration: How irritated, stressed, and annoyed versus content, relaxed, and complacent did you feel during the task?



Move 1 block

Page Break

 Import from library

Add new question

Add Block

▼ System Usability Survey

Q9

System Usability: Please read each statement and select your level of agreement.

Q10



I think that I would like to use this system frequently

Strongly disagree

Strongly agree

☐☐☐☐☐

Q11



I found the system unnecessarily complex

Strongly disagree

Strongly agree

☐☐☐☐☐

Q12



I thought the system was easy to use

Strongly disagree

☐☐☐☐

Strongly agree

☐

Q13



I think that I would need the support of a technical person to be able to use this system

Strongly disagree

☐☐☐☐

Strongly agree

☐

Move 1 block

Q14



I found the various functions in this system were well integrated

Strongly disagree

☐☐☐☐

Strongly agree

☐

Q15



I thought there was too much inconsistency in this system

Strongly disagree

☐☐☐☐

Strongly agree

☐

Q16



I would imagine that most people would learn to use this system very quickly

Strongly disagree

☐☐☐☐

Strongly agree

☐

Q17



I found the system very cumbersome to use

Strongly disagree

☐☐☐☐

Strongly agree

☐

Q18



I felt very confident using the system

Strongly disagree

☐☐☐☐

Strongly agree

☐

Q19



I needed to learn a lot of things before I could get going with this system

Strongly disagree

☐☐☐☐

Strongly agree

☐

Move 1 block

 Import from library

Add new question

Add Block

▼ Presence

Q20

Presence: Please read each question and choose the appropriate level.

Q21



How much were you able to control events?

Not at all

☐☐☐

Moderately

☐☐☐

Completely

☐

Q22



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

Not at all

☐☐☐

Moderately

☐☐☐

Completely

☐

Q23



How natural did your interactions with the environment seem?

Not at all

Moderately

Completely

☐☐☐☐☐☐☐

Q24



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

Not at all

Moderately

Completely

☐☐☐☐☐☐☐

Move 1 block

Q25



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

Not at all

Moderately

Completely

☐☐☐☐☐☐☐

Q26



How compelling was your sense of objects moving through space?

Not at all

Moderately

Completely

☐☐☐☐☐☐☐

Q27



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

Not at all

Moderately

Completely

☐☐☐☐☐☐☐

Q28



Were you able to anticipate what would happen next in response to the actions that you performed?

Not at all

Moderately

Completely

☐☐☐☐☐☐☐

Q29



How completely were you able to actively survey or search the environment using vision?

Not at all

☐☐☐

Moderately

☐☐☐

Completely

☐

Q30



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

Not at all

☐☐☐

Moderately

☐☐☐

Completely

☐

Q31



How well could you move or manipulate objects in the virtual environment?

Not at all

☐☐☐

Moderately

☐☐☐

Completely

☐

Q32



How involved were you in the virtual environment experience?

Not at all

☐☐☐

Moderately

☐☐☐

Completely

☐

Q33



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

Not at all

☐☐☐

Moderately

☐☐☐

Completely

☐

Q34



How quickly did you adjust to the virtual environment experience?

Not at all

☐☐☐

Moderately

☐☐☐

Completely

☐

Q35



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

Not at all

☐☐☐

Moderately

☐☐☐

Completely

☐

Q36



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

Not at all

☐☐☐

Moderately

☐☐☐

Completely

☐

Q37



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

Not at all

☐☐☐

Moderately

☐☐☐

Completely

☐

Q38



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

Not at all

☐☐☐

Moderately

☐☐☐

Completely

☐

Q39



How completely were your senses engaged in this experience?

Not at all

☐☐☐

Moderately

☐☐☐

Completely

☐

Q40



How easy was it to identify objects through physical interaction, like touching an object, walking over a surface, or bumping into a wall or object?

Not at all

☐☐☐

Moderately

☐☐☐

Completely

☐

Q41



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

Not at all

Moderately

Completely



Q42



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

Not at all

Moderately

Completely

Move 1 block



Q43



Was the information provided through different senses in the virtual environment (e.g., vision, hearing, touch) consistent?

Not at all

Moderately

Completely



Import from library

Add new question

Add Block

▼ Affordances in VR

Q44



Affordances in VR: Please read each statement and rate the value of each listed feature.

Least valuable

Most valuable



Q45



Working in VR, with respect to understanding the problem, I found the ability to navigate from call graph to control flow graph to code:

Least valuable

Most valuable



Q46



Working in VR, with respect to being able to recall information, I found being able to place objects anywhere in the environment:

Least valuable

Most valuable

☐☐☐☐☐

Q47



Working in VR, with respect to being able to recall information, I found viewing individual function call graphs and control flow graphs:

Move 1 block

Least valuable

Most valuable

☐☐☐☐☐

Q48



Working in VR, with respect to formulating a solution to the problem, I found the visual callouts of function capabilities:

Least valuable

Most valuable

☐☐☐☐☐

Q49



Working in VR, with respect to formulating a solution to the problem, I found the text-based information (such as disassembly, string listings, etc.):

Least valuable

Most valuable

☐☐☐☐☐

Q50



Working in VR, with respect to formulating a solution to the problem, I found being able to place objects anywhere in the environment:

Least valuable

Most valuable

☐☐☐☐☐

 Import from library

Add new question

Add Block

End of Survey

We thank you for your time spent taking this survey.

Your response has been recorded.

Revised 07/12/2022

Appendix D (Health info)

VR Screening Questions

To be provided to participants before starting a virtual reality-based session; adapted from [10]:

Review the questions below and answer them silently.

1. Do you have any cold or flu symptoms?
2. Do you have any open cuts or sores on your face or hands?
3. Do you have an empty stomach (have not eaten today)?
4. Do you have any other pre-existing medical conditions that would prevent effective use of the virtual reality equipment? (e.g., hand pain or deformity, strabismus (crossed eyes) or amblyopia (lazy eye), etc.)

If “yes” to any of the above, we unfortunately recommend you do not continue with this study.

5. Do you have any other pre-existing medical conditions that may impact the quality of your experience with the virtual reality equipment? (e.g., epileptic; pregnant; suffer from motion sickness or balance problems; susceptible to migraines; etc.)
6. Have you experienced eye strain?
7. Have you experienced stress induced anxiety?

If “yes” to any of the above, proceed with caution and let the facilitator know if you are experiencing any symptoms during the virtual reality experience.

The Auburn University Institutional
Review Board has approved this
Document for use from
07/23/2024 to -----
Protocol # 24-668 EP 2407



This is to certify that:

Dennis Brown

Has completed the following CITI Program course:

IRB # 2 Social and Behavioral Emphasis - AU Personnel - Basic/Refresher

(Curriculum Group)

IRB # 2 Social and Behavioral Emphasis - AU Personnel

(Course Learner Group)

1 - Basic Course

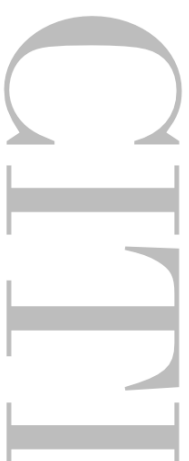
(Stage)

Under requirements set by:

Auburn University

Completion Date 06-Nov-2021
Expiration Date 05-Nov-2024
Record ID 45930836

Not valid for renewal of
certification through CME.



Collaborative Institutional Training Initiative

101 NE 3rd Avenue, Suite 320
Fort Lauderdale, FL 33301 US
www.citiprogram.org

Generated on 23-Jun-2024. Verify at www.citiprogram.org/verify/?w7a1bb671-cb24-4b7f-91c0-099b83336a88-45930836



This is to certify that:

Dennis Brown

Has completed the following CITI Program course:

Responsible Conduct of Research

(Curriculum Group)

AU Basic RCR Training for ALL Faculty, Staff, Postdocs, and Students

(Course Learner Group)

1 - RCR

(Stage)

Under requirements set by:

Auburn University

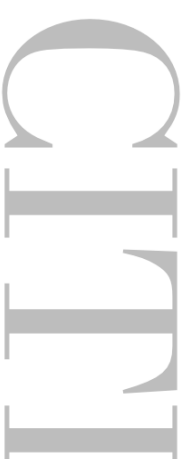


Completion Date 03-Feb-2023

Expiration Date 03-Feb-2026

Record ID 54188111

Not valid for renewal of
certification through CME.



Collaborative Institutional Training Initiative

Verify at www.citiiprogram.org/verify/?wdd15d1cf-1714-4b57-90c9-7b61eb9ba1ed-54188111

[EXT] CITI Program - Course Completion for Samuel Mulder

CITIPROGRAM No Reply<noreply@citiiprogram.org>
To: Samuel Mulder

Mon 7/1/2024 5:32 PM

CAUTION: Email Originated Outside of Auburn.



Course Completion for Samuel Mulder

Congratulations on your recent course completion!

Name:	Samuel Mulder (ID: 10785729)
Institution:	Auburn University (ID: 964)
Course:	IRB # 2 Social and Behavioral Emphasis - AU Personnel
Stage:	1 - Basic Course
Completion Date:	01 Jul 2024
Expiration Date:	01 Jul 2027
Completion Record ID:	63514492

To share the **Completion Report** for this course, use the following link:

citiiprogram.org/verify/?kfd68934-88e1-4c0f-9879-7b6d717c3959-63514492

Note that this link will share the full two-part report, which includes all quiz scores.



This is to certify that:

Kevan Baker

Has completed the following CITI Program course:

IRB # 2 Social and Behavioral Emphasis - AU Personnel - Basic/Refresher

(Curriculum Group)

IRB # 2 Social and Behavioral Emphasis - AU Personnel

(Course Learner Group)

1 - Basic Course

(Stage)

Under requirements set by:

Auburn University



Collaborative Institutional Training Initiative

101 NE 3rd Avenue, Suite 320

Fort Lauderdale, FL 33301 US

www.citiprogram.org

Completion Date 24-Jun-2024

Expiration Date 24-Jun-2027

Record ID 63546875

Not valid for renewal of
certification through CME.

Generated on 24-Jun-2024. Verify at www.citiprogram.org/verify/?wb500ae7b-e470-4b20-a441-22e73047a606-63546875



This is to certify that:

Kevan Baker

Has completed the following CITI Program course:

Responsible Conduct of Research

(Curriculum Group)

AU Basic RCR Training for ALL Faculty, Staff, Postdocs, and Students

(Course Learner Group)

1 - RCR

(Stage)

Under requirements set by:

Auburn University

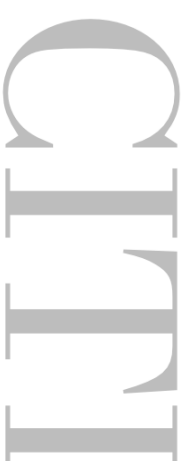


Completion Date 02-Jul-2024

Expiration Date 02-Jul-2027

Record ID 63701696

Not valid for renewal of
certification through CME.



Collaborative Institutional Training Initiative

101 NE 3rd Avenue, Suite 320

Fort Lauderdale, FL 33301 US

www.citiprogram.org

Generated on 02-Jul-2024. Verify at www.citiprogram.org/verify/?we8a43064-f793-4a9f-94f6-03ba0c4b7bbf-63701696



This is to certify that:

Julian Bauer

Has completed the following CITI Program course:

IRB # 2 Social and Behavioral Emphasis - AU Personnel - Basic/Refresher

(Curriculum Group)

IRB # 2 Social and Behavioral Emphasis - AU Personnel

(Course Learner Group)

1 - Basic Course

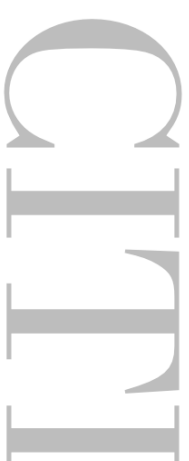
(Stage)

Under requirements set by:

Auburn University

Completion Date 26-Jun-2024
Expiration Date 26-Jun-2027
Record ID 63571827

Not valid for renewal of
certification through CME.



Collaborative Institutional Training Initiative

101 NE 3rd Avenue, Suite 320
Fort Lauderdale, FL 33301 US
www.citiprogram.org

Generated on 26-Jun-2024. Verify at www.citiprogram.org/verify/?w4970fb01-9c87-4979-96b2-5a8c08386723-63571827



This is to certify that:

Julian Bauer

Has completed the following CITI Program course:

Responsible Conduct of Research

(Curriculum Group)

AU Basic RCR Training for ALL Faculty, Staff, Postdocs, and Students

(Course Learner Group)

1 - RCR

(Stage)

Under requirements set by:

Auburn University

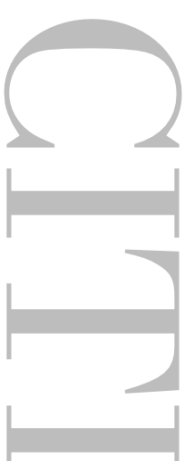


Completion Date 03-Jul-2024

Expiration Date 03-Jul-2027

Record ID 63717549

Not valid for renewal of
certification through CME.



Collaborative Institutional Training Initiative

101 NE 3rd Avenue, Suite 320

Fort Lauderdale, FL 33301 US

www.citiprogram.org

Generated on 03-Jul-2024. Verify at www.citiprogram.org/verify/?w37118540-b10e-40ef-bcc4-017109c70ffe-63717549

Delete Archive Report Reply Reply all Forward Zoom

VENDOR VETTING RESULTS: HTC America Inc. - HTC VIVE Pro 2

A Auburn University Vendor Vetting <noreply@workflownotification.com> Reply Reply all Forward
To: Dennis Brown Mon 7/22/2024 5:15 PM
Cc: Shannon Price; Barbara Cosby; Dudley Dent; Mark Bransby; Nathan Waters

Some content in this message has been blocked because the sender isn't in your Safe senders list.

[Trust sender](#)

[Show blocked content](#)

Dennis Brown,

Manufacturer: HTC America Inc.

Product: HTC VIVE Pro 2

Decision Date: 7/22/2024

Expiration Date: 7/22/2027 5:14:52 PM

Cybersecurity Note (if any): VendorVetting - AUSEcurity (Approve) 7/22/2024 1:02:25 PM - 7/22/2024 1:09:20 PM - [HTC VIVE Pro 2 has been approved for purchase by Cybersecurity. Care must be taken to operate the application or system securely. According to the Cybersecurity Roles and Responsibilities Policy, "Application Owners (colleges, departments or individual users) are responsible for establishing, maintaining, and documenting security controls for all IT systems and data they control."

Please let Cybersecurity know if we can be of assistance in helping you implement security for this request.]

Status: **Approved**

Print this email to PDF and include it along with other required documentation in the eRequisition system for Purchasing and Business Services.

*Notes: *If someone on the team, current or future, will be prevented from using the hardware and/or software because of a physical disability, an Equal Effective Alter*

Send an email to AU Vendor Vetting at vetting@auburn.edu if you have any questions.

Thank you,
AU Vendor Vetting

Auburn University Logo

Reply

Reply all

Forward