Assessing The affect of Predictability on Interaction With AI in VR

Unveiling Insights into Human-AI Interaction Dynamics

Master’s thesis in Computer science and engineering

ISHWOR KARKI

ISMAEL ALBUTIHE

Department of Computer Science and Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
UNIVERSITY OF GOTHENBURG
Gothenburg, Sweden 2024

Master's thesis 2024

Assessing The affect of Predictability on Interaction With AI in VR

Insights into Human-AI Interaction Dynamics

ISHWOR KARKI

ISMAEL ALBUTIHE

Department of Computer Science and Engineering
Chalmers University of Technology
University of Gothenburg
Gothenburg, Sweden 2024

Assessing The affect of Predictability on Interaction With AI in VR
Insights into Human-AI Interaction Dynamics
ISHWOR KARKI AND ISMAEL ALBUTIHE

Supervisor: Palle Dahlstedt. Professor Interaction Design (IxD)
Supervisor: Paweł W. Woźniak, Computer Science and Engineering
Examiner: Staffan, Björk

Master's Thesis 2024
Department of Computer Science and Engineering
Chalmers University of Technology and University of Gothenburg
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Typeset in LaTeX
Gothenburg, Sweden 2024

Assessing The affect of Predictability on Interaction With AI in VR
Insights into Human-AI Interaction Dynamics
ISHWOR KARKI & ISMAEL ALBUTIHE
Department of Computer Science and Engineering
Chalmers University of Technology and University of Gothenburg

Abstract

In this era of technological innovation, the fusion of artificial intelligence (AI) with user experience (UX) presents a transformative shift in human-computer interaction (HCI). This thesis investigates the impact of predictability in AI-driven virtual reality (VR) environments on user experience and task performance. The pre-study aimed to operationalize the concept of "predictability," revealing significant differences between predictable, unpredictable, and manual conditions. The main study focused on task completion duration, user satisfaction, perceived control, cognitive load, and error rates across these conditions. Results indicated that participants completed tasks significantly faster in the predictable condition compared to the manual and unpredictable conditions, suggesting that predictability enhances efficiency and streamlines user interactions. However, no significant differences were found in user satisfaction, perceived control, NASA Task Load Index (NASA-TLX) scores, or error rates across conditions. Qualitative feedback revealed that while the predictable condition was described as "smooth" and "satisfying," the unpredictable condition elicited frustration and confusion, highlighting the importance of considering both quantitative and qualitative data in user experience evaluation. The study's limitations include the novelty effect of VR and the focus on a specific task, which may not fully capture real-world AI interactions. Future research should explore predictability in diverse AI applications, conduct longitudinal studies, and consider user diversity to enhance the generalizability of findings. This study underscores the critical role of predictability in AI systems, providing valuable insights for designing more intuitive and efficient AI-driven environments.

Keywords: Computer science, Engineering, Predictability, Artificial Intelligence (AI), Human–AI Interaction, Joint Action..

Acknowledgements

First and foremost, we would like to express our sincere gratitude to our supervisor Paweł W. Woźniak, for his invaluable guidance, support, and patience throughout the entirety of this thesis. His expertise and encouragement have been instrumental in shaping our research and academic journey.
We would also like to extend our thanks to Professor Palle Dahlstedt, who provided crucial assistance in finalizing our work during Paweł's absence. His insights and support were invaluable in bringing our thesis to completion.
We extend our heartfelt thanks to the faculty and staff at Chalmers University for providing us with the resources and environment conducive to scholarly inquiry. Their dedication to fostering intellectual curiosity and academic excellence has been truly inspiring.
Additionally, we would like to acknowledge our partners and family for their unwavering support and understanding, which has been a source of strength and motivation during challenging times. Their encouragement has played a pivotal role in our academic pursuits.
Finally, we express our gratitude to all individuals who have engaged with us during the course of this thesis, whether through discussions, feedback, or support. Your contributions have enriched our learning experience and facilitated our growth in myriad ways.

Ishwor Karki and Ismael Albutihe, Gothenburg, May 2024

Contents

List of Figures **xiii**

List of Tables **xiv**

1 Introduction **1**

1.1 Aim and Purpose 2

1.2 Objectives 2

1.3 Thesis Overview 3

2 Theory & Related Work **4**

2.1 Predictability 4

2.2 Joint Actions 5

2.3 Automation & AI 6

2.4 Preeceding Research and Contrasts 7

3 Methodology **11**

3.1 Research Method 11

3.1.1 Potential Research Methods 11

3.1.1.1 Qualitative Methods 12

3.1.1.2 Quantitative Methods 12

3.1.1.3 Mixed-Methods Approach 12

3.1.2 Feasibility of Potential Methods 13

3.1.3 Chosen Research Methods 13

3.1.4 Rationale for Choosing Mixed-Methods Approach 14

3.2 System/Software 14

3.2.1 Prototyping 14

3.2.1.1 Virtual Reality 15

3.2.1.2 Unity 15

3.2.1.3 Blender 15

3.2.2 Data Collection 16

3.2.2.1 PsyToolkit 17
3.2.2.2 Prolific.com 17
3.2.2.3 Flutter 17
3.2.3 Analysis and Data Visualization 18
3.2.3.1 R Studio 18
3.2.3.2 Atlas.ti 19
3.2.4 Roles and Responsibilities in Thesis Development 21

4 Method **22**
4.1 Pre-Study 22
4.1.1 Study Design 23
4.1.2 Participants 23
4.1.3 Materials and Instruments 23
4.1.4 Procedure 26
4.1.5 Data Analysis 26
4.1.6 Results 27
4.1.6.1 Descriptive Statistics 27
4.1.7 Quantitative Results 27
4.1.8 Discussion 28
4.2 Pilot-Study 29
4.2.1 Study Design 29
4.2.2 Participants 30
4.2.3 Materials and Instruments 30
4.2.4 Procedure 30
4.2.5 Data Analysis 30
4.2.6 Outcomes and Modifications 31
4.2.7 Discussion 31
4.3 Main Study 31
4.3.1 Study Design 32
4.3.2 Participants 32
4.3.3 Materials and Instruments 32
4.3.4 Procedure 32
4.3.5 Data Analysis 33
4.3.6 Outcomes and Modifications 33
4.3.7 Results 33
4.3.7.1 Descriptive Statistics 34
4.3.7.2 Quantitative Results 35
4.3.7.2.1 Control 35

4.3.7.2.2 Duration 35
4.3.7.2.3 Error Rate 37
4.3.7.2.4 NASA Task Load Index 38
4.3.7.2.5 SUS 39
4.3.7.3 Qualitative Results 39
4.3.7.3.1 Manual 39
4.3.7.3.2 Predictable AI 40
4.3.7.3.3 Unpredictable AI 41

5 Discussion **42**
5.1 Task Completion Duration 42
5.2 User Satisfaction 42
5.3 Control 43
5.4 Task Load 44
5.5 Error Rate 44
5.6 Realistic Expectations for AI Capabilities 45
5.7 Influence of Demographics and the Age Factor 46
5.8 Other Factors Affecting User Experience 46
5.9 Qualitative Insights 47
5.10 Applicability of Findings to Other Settings, Such as Autonomous Driving 47
5.11 VR Experience 48
5.12 Ethical Considerations 48
5.13 Limitations 48
5.14 Future Research 49

6 Conclusion **50**

A Study Design Document **I**
A.1 Purpose of study (research hypothesis/questions) I
A.2 Participant profile I
A.3 Compensation plan I
A.4 Methodology I
A.4.1 Study Design I
A.4.2 Experiments II
A.4.3 Variables II
A.5 Material & Apparatus II
A.6 Structure & Timeline III

A.6.1 Preparation required before the study III
A.6.2 Timeline of study phase III
A.7 Analysis IV
A.7.1 Quantitative Analysis IV
A.7.2 Qualitative Analysis IV
A.7.3 Data Forms/Logs IV
A.8 Personnel Involved & Responsibilities IV

B Consent Form **V**

C Nasa-TLX Form **VII**

List of Figures

3.1 Prototype designed using blender and unity 16
3.2 Designing the survey using PsyToolKit 17
3.3 Some images of the Web App taken during the experiment 18
3.4 Analyzing the result from experiment using R programming 19
3.5 Qualitative analysis map using Atlas.ti 20

4.1 2x2 Grid of participant performing under manual condition 24
4.2 2x2 Grid of participant performing under predictable condition 25
4.3 2x2 Grid of participant performing under unpredictable condition . . 26
4.4 Box Plot in R of Predictability under three conditions 28
4.5 Box Plot in R of Control under three conditions 35
4.6 Box Plot in R of Time Duration under three conditions 36
4.7 Box Plot in R of Error Rate under three conditions 37
4.8 Box Plot in R of NASA-TLX under three conditions 38
4.9 Box Plot in R of SUS under three conditions 39

C.1 NASA-TLX VII

List of Tables

3.1 Thesis Work Contribution 21

4.1 Descriptive Data from Survey 27
4.2 Descriptive Data from Experiment 34

1 Introduction

In the dynamic age of Artificial Intelligence (AI) advancements, there's a growing interest in applying AI technologies to refine User Experience (UX) and streamline Human-Computer Interaction (HCI) (Yang et al., 2020). This excitement stems from AI's ability to profoundly reshape various facets of HCI, presenting enhanced user experiences across a spectrum of applications. Ranging from basic utilities, such as spam filters, to complex endeavors like autonomous vehicles, the contributions of AI are becoming increasingly crucial in HCI research. This trend has sparked a wave of interest among designers and researchers in developing innovative AI applications capable of significantly transforming user interactions.

Recent developments in AI tools, such as GPT-3 for natural language processing and DALL · E for image generation, underscore the significant strides made in AI technologies. These tools not only automate tasks that were once considered the exclusive purview of human intellect but also pave the way for creating more intuitive and engaging user interfaces. However, the integration of AI into Human-Computer Interaction (HCI) introduces unique challenges. Designing interactions with AI systems can become intricate, especially when users find it difficult to predict the system's behavior, potentially leading to usability concerns. A recent study highlighted the issue of unpredictable technology responses and the resulting "creepiness," (Woźniak et al., 2021). This topic is not widely explored in Human-AI interactions. This situation demands a thoughtful approach to design interactions that are adaptable to AI's dynamic nature, focusing on predictability and user comfort. Emphasizing these aspects is crucial to maximizing AI's potential to improve the user experience.

To address these intricacies, our study uses Virtual Reality (VR) as a methodological tool to assess predictability and its effects on Human-AI interaction. VR enables the simulation of various interactions in controlled, immersive environments, facilitating the close observation and measurement of user responses to different technological behaviors (Weistroffer et al., 2013). The immersive nature of VR provides a unique and effective platform for conducting studies that offer deeper insights than those typically achieved through traditional ones, thus enhancing our understanding of optimal design and implementation strategies for technology use (Weistroffer et al., 2013).

Meeting these challenges necessitates an exploration of predictability and its impact on Human-AI interaction. This study is dedicated to examining these elements in detail, with the goal of uncovering how predictability can guide the design and implementation of AI systems in a manner that not only elevates the user experience but also maintains a high degree of user confidence and satisfaction.

1.1 Aim and Purpose

The aim of this study is to investigate the influence of user predictability on Human-Artificial Intelligence (AI) interaction within Virtual Reality (VR) environments. Through a combination of survey-based operationalization and experimental investigation, this research aims to uncover the nuanced relationship between user predictability and satisfaction, shedding light on the underlying factors shaping user experiences in VR-based AI interactions. By elucidating the impact of predictability on user satisfaction, this study seeks to provide valuable insights for designing more intuitive and engaging Human-AI interaction systems in VR, thus contributing to the advancement of user experience design and HCI research in emerging technological domains.
The purpose of this study is to answer the following research question: "**To what extent does the predictability of AI behavior affect user satisfaction during object manipulation tasks in Human-AI interaction within Virtual Reality environments?**"

1.2 Objectives

To fulfill the objectives of this study, the following steps will be undertaken: Firstly, conduct a survey involving 27 participants to operationalize the concept of user predictability and gather initial insights into user expectations and preferences regarding Human-AI interaction within Virtual Reality (VR) environments. Secondly, carry out an experimental study with 21 participants to further investigate the impact of user predictability on user satisfaction in Human-AI interaction within VR settings, utilizing controlled conditions to assess user responses to varying levels of predictability. Thirdly, analyze the data collected from both the survey and experimental study to clearify the relationship between user predictability and satisfaction, as well as to identify factors influencing user perceptions and behaviors in Human-AI interaction scenarios within VR. Finally, synthesize the findings to develop actionable recommendations for enhancing user predictability and satisfaction in Human-AI interaction design for VR applications, contributing to the advancement of understanding and practice in this evolving field.

1.3 Thesis Overview

This study anticipates the following outcomes: 1) Survey findings are expected to reveal varying user preferences regarding predictability in VR-based Human-AI interaction, potentially indicating a preference for higher predictability levels. 2) Experimental results are projected to demonstrate a significant correlation between user predictability and satisfaction, suggesting that higher predictability enhances user satisfaction. 3) User behavior analysis is expected to show that user engagement, task completion rates, and emotional responses vary with the level of predictability, possibly indicating increased confidence with higher predictability. 4) The study aims to identify key design principles for enhancing user predictability and satisfaction in VR-based Human-AI interaction, offering actionable guidelines for developers. 5) Overall, the study aims to contribute empirical evidence on the importance of predictability in shaping user satisfaction, providing valuable insights for HCI researchers and practitioners in AI technology development for VR environments.

2

Theory & Related Work

2.1 Predictability

Predictability plays a pivotal role in human cognition, as extensively explored in the field of cognitive neuroscience. Bubic et al. (2010) have significantly contributed to our understanding by highlighting how predictive processing is intricately woven into various cognitive domains. Their research reveals that prediction transcends mere anticipation of future events; it is a fundamental aspect of cognitive and neural functioning that enhances perceptual systems, optimizes decision-making processes, and refines motor actions through anticipatory adjustments. The brain employs predictive mechanisms to effectively bridge past experiences with future actions, facilitating more efficient navigation and interaction with the environment. These processes are not only crucial in human cognition but also demonstrate rudimentary forms in animal behavior, highlighting their evolutionary significance. By integrating predictive processing, cognitive frameworks transition from traditional reactive models to ones where cognition is primarily anticipatory, continually projecting forward in time to prepare for upcoming events. This paradigm shift underscores the profound implications of predictability for the fundamental structuring of cognitive processes and their development.

Predictability also emerges as a critical factor in human-AI or human-robot interactions, particularly within the context of artificial agents such as software algorithms, robots, and AI systems. The predictability of these agents significantly impacts human trust, reliance, and interaction satisfaction (Hoff & Bashir, 2015; Lee & See, 2004). For example, predictable behavior from nonhuman agents has been shown to facilitate smoother interactions, allowing users to anticipate responses and effectively adjust their actions (J. Johnson, 2020). This is particularly crucial in contexts where safety and efficiency are paramount, such as automated driving systems or collaborative robotics (Kim & Hinds, 2006). Moreover, predictable algorithms not only help in building initial trust but are essential for fostering long-term user engagement and adaptation (Madhavan & Wiegmann, 2007). By enhancing user comfort and confidence, predictability in nonhuman agents makes the interaction experience more intuitive and rewarding, thereby underscoring the necessity of incorporating predictability into the design of cognitive systems, establishing it as a foundational element that shapes the dynamics of interaction across various technological domains.

Enhancing user comfort and confidence can be closely linked with usability testing, a commonly utilized method for identifying usability issues (Kuang, 2023). Usability encompasses various elements such as the consistency and ease with which users can manipulate and navigate a website, the clarity of interaction, readability, information arrangement, speed, and layout (de Vreede et al., 2005). Assessing the usability of a system involves considering predictability or unpredictability as one aspect, determining how predictable a specific task is (Fernando & Hodrien, 2021).

2.2 Joint Actions

Joint action refers to the coordination between two or more individuals in space and time to achieve a change in the environment. This concept is crucial in understanding both human-human and human-robot (or human-AI) interactions. In the context of human-human joint action, it involves the coordination of activities among individuals to perform tasks that result in a shared outcome, where both participants form a 'we' identity at a pre reflective level, even if their subjective experiences of agency and contributions to the outcome may differ (Obhi & Hall, 2011; Sebanz et al., 2006; Strother et al., 2010). Human-robot or human-AI joint action extends this concept to interactions where one of the agents is an automated tool or robot, introducing complexities in the understanding and coordination of actions between human and non-human actors. The study of joint action, especially in human-AI interactions, is important because it aids in the development of frameworks for effective human-AI collaboration. Such studies help in designing autonomous tools that can interact seamlessly with human partners, considering aspects like perception, decision-making, action, communication, and learning, as well as their integration (Clodic et al., 2017). The exploration of joint action across these domains is crucial for advancing our understanding of social interaction mechanisms and enhancing the design and functionality of AI and robotic systems for improved human-AI collaboration.

2.3 Automation & AI

Automation represents a significant shift in technology, transitioning from basic mechanical processes to complex systems that incorporate artificial intelligence (AI). These systems are designed to mimic human cognitive functions, including learning, decision-making, and problem-solving. As these technologies advance towards autonomy, human roles have shifted from passive observers to active participants in this dynamic interaction (Groover, 2020).
The infusion of AI into automation, evident in Artificial Intelligence Systems (AIS), remote-control devices, and cutting-edge innovations like self-driving cars and robots, signifies a notable advancement in engineering and technology. This progression facilitates daily interactions with advanced automated systems, aiming to alleviate cognitive and physical strains while boosting human capabilities. Consequently, this fosters a new era of cooperation, referred to as "Human-Robot Joint Action" or "Human–AI Interaction," which reflects the evolving dynamics of our relationship with these technologies (Wen & Haggard, 2018; Wen & Imamizu, 2022).
In *Artificial Intelligence: A Modern Approach*, Russell and Norvig (2021) categorize AI into various types based on how they function and process information. One foundational approach is rule-based AI, also known as symbolic AI. Rule-based AI relies on predefined rules and formal logic to represent knowledge and make decisions. These systems apply logical rules to input data, often using "if-then" structures to derive conclusions. While rule-based AI is transparent and interpretable, making it ideal for tasks such as legal reasoning and diagnostic tools, it struggles in environments where rules become difficult to define, particularly in complex or unstructured situations. This limitation led to the development of more adaptive AI methods.
In contrast, other AI approaches like machine learning (ML) and neural networks enable systems to learn from data rather than rely solely on pre-programmed rules. ML algorithms, including deep learning models, have become essential in modern AI applications such as image and speech recognition. Additionally, search-based AI and optimization-based AI explore possible solutions within large search spaces or seek the best solutions under certain constraints. Bayesian networks and probabilistic AI handle uncertainty by modeling probabilistic relationships between variables, offering robustness in decision-making. Evolutionary algorithms, such as genetic algorithms, further enhance AI's ability to evolve solutions over time by mimicking natural selection processes. Lastly, hybrid AI systems, which combine multiple approaches, have become increasingly common to leverage the strengths of various methodologies. These diverse techniques demonstrate the breadth of AI, from struc-

tured, logic-based rule systems to the learning and adaptability seen in modern AI methods (Russell & Norvig, 2020)
Tracing back to the start of AI in 1956 with the creation of the first AI program (Newell & Simon, 1956), there has been a consistent effort to mimic human cognition and develop entities capable of autonomous decision-making (Russell & Norvig, 2020). This progression towards advanced automation, equipped with a variety of AI functionalities, highlights our ongoing dedication to building machines that can perceive, understand, predict, and navigate complex scenarios in ways that resemble human intelligence. Therefore, enhancing human-AI interaction remains a paramount goal, crucial for shaping the future of automation and its seamless integration into our lives.

2.4 Preeceding Research and Contrasts

In recent decades, there has been a significant amount of research conducted in the field of human-AI interactions. With AI at the cutting edge, it has fascinated many. Our goal, however, was to study the effects of being able to predict how AI-driven systems would behave. Here are some related works that we reviewed.

Yang et al. (2020) examined the distinctive challenges posed by designing human-AI interactions, emphasizing two critical aspects: Capability Uncertainty and Output Complexity. Their study identifies how the unpredictability of AI systems complicates the design process, as designers struggle to anticipate how AI capabilities will evolve and perform in real-world applications. These challenges are particularly pronounced in systems that adapt and learn over time, making it difficult to envision and prototype accurate AI behaviors and their potential errors. By mapping the difficulties AI introduces into the design process, the research highlights that traditional methods—such as sketching and prototyping—often fail to accommodate AI's complexity. The study further illustrates how these issues impact user experience, particularly when designing AI interactions that rely on evolving, adaptive outputs. This work contributes to a better understanding of how to approach designing for predictability in AI systems, offering insights for developing more robust frameworks for human-AI collaboration and addressing the inherent unpredictability in such systems.

Rzepka and Berger (2018) explores user interactions with various AI-enabled systems, such as expert systems, chatbots, recommender systems, and robots, using a

Human-Computer Interaction (HCI) framework. It highlights how AI systems incorporate advanced capabilities like learning, autonomy, and natural language processing, which influence user perceptions and behaviors. Users often assign human-like traits to AI systems, leading to both positive responses, such as increased trust, and negative ones, like threat perceptions, especially with highly autonomous systems. The paper emphasizes the importance of system transparency and human-like behaviors, which positively affect user trust and acceptance. Despite growing research, gaps remain in understanding how different AI capabilities and levels of transparency impact user interaction, suggesting a need for future studies on newer AI systems, such as autonomous vehicles and voice assistants. This review consolidates dispersed findings, offering a comprehensive foundation for studying AI-user interactions in IS research.

Bergström et al. (2021) reviews two decades of research on object selection and manipulation in virtual reality (VR) and highlights the lack of standardized guidelines for evaluating these interactions. While various techniques have been developed, existing 2D frameworks like Fitts' law are inadequate for VR due to its unique spatial and physical interaction requirements. The authors analyze best practices from previous studies and propose design guidelines and a reporting checklist to improve consistency, replicability, and ecological validity in VR research. They emphasize the need for more realistic study designs that address VR-specific factors such as depth perception, occlusion, and user discomfort. The review underscores the importance of establishing standardized evaluation methods to enable better comparisons and knowledge accumulation across studies, with the goal of improving the rigor and validity of future VR experiments on object selection and manipulation.

Boffetta et al. (2002) conducted a comprehensive review of the predictability problem in dynamical systems, exploring the intricate relationships among Lyapunov exponents, Kolmogorov-Sinai entropy, Shannon entropy, and algorithmic complexity. This study highlights how characterizing the unpredictability of a system can provide a measure of its complexity. Further research by Boffetta et al. (2002) reviews various developments in characterizing predictability across systems exhibiting different types of complexity, from low-dimensional systems to high-dimensional systems with spatio-temporal chaos and fully developed turbulence. Special attention is given to the effects of finite-time and finite-resolution on predictability, proposing suitable generalizations of standard indicators to account for these effects. Additionally, the study addresses the challenges involved in systems with intrinsic randomness, focusing on distinguishing chaos from noise and modeling the system. The characterization of irregular behavior in systems with discrete phase space is also examined.

Li et al. (2023) developed a research protocol for evaluating human-AI interaction within specific AI products, aiming to assist UX and HCI researchers in assessing various interaction solutions and validating design decisions prior to engineering investments. The study provides a comprehensive description of the research protocol and demonstrates its application by examining an existing set of human-AI interaction guidelines. Using factorial surveys with a 2×2 mixed design, the study compared user perceptions when guidelines were followed versus violated, under conditions of both optimal and sub-optimal AI performance. The findings offered both qualitative and quantitative insights into the UX impact of each guideline. These insights are intended to aid designers of user-facing AI systems in prioritizing and applying the guidelines more effectively. Although no new AI systems were developed in this thesis, it offers a valuable methodology for evaluating and refining human-AI interaction strategies.

Swan and Notess (2003) conducted an ongoing study comparing user satisfaction ratings obtained from user tests with those collected following actual use of digital music library software. The study identifies variables that hinder accurate prediction and evaluates the utility of surveys in predicting satisfaction gaps between test subjects and real-world users. Using satisfaction questionnaires, the research compares baseline user satisfaction with an existing software version, Variations, against both test subject and user satisfaction with a new version, Variations2. The study involved three parts: a baseline satisfaction survey with 30 music students using Variations, a usability test with 10 subjects for Variations2, and a real-world usage survey with 12 students using Variations2 for a course assignment. Results indicated a lower mean satisfaction rating for Variations2 compared to Variations, and a notable difference between test subjects and real-world users' satisfaction with Variations2. This work aims to explore whether satisfaction data from user tests can reliably predict real-world satisfaction, providing insights into the practical value of such measurements. Although no definitive conclusions are drawn, this research underscores the importance of evaluating and improving the predictive power of user satisfaction surveys in the context of software usability.

In our study, we are focusing on investigating the impact of predictability within AI-driven virtual reality (VR) environments on user experience and task performance. Unlike Li et al. (2023) research protocol, which evaluates human-AI interaction within specific AI products, and Swan and Notess (2003) study, which examines the reliability of satisfaction data from user tests in predicting real-world satisfaction with digital music library software, our research uniquely explores the dynamics of user interactions within immersive VR settings. Additionally, while Boffetta et al. (2002) study delves into the predictability problem in dynamical systems, our study extends beyond theoretical complexity measures to empirically investigate how predictability influences user satisfaction, perceived control, cognitive load, and error rates in AI-driven VR environments. Through this approach, we aim to contribute valuable insights for designing more intuitive and engaging Human-AI interaction systems in VR contexts.

3

Methodology

This chapter outlines the research methodology used to investigate the effect of AI-driven predictability on user experience and task performance in Virtual Reality (VR) environments.

3.1 Research Method

In this section, we present the research methodology employed in this study, designed to explore the impact of AI-driven predictability on user experience and task performance within Virtual Reality (VR) environments. To answer our research question, we adopted a mixed-methods approach, integrating both qualitative and quantitative data collection techniques. This approach was chosen to capture the multifaceted nature of human-AI interactions, where objective performance metrics alone may not fully reflect user experiences or cognitive responses (Clark & Maguire, 2020).

We begin by outlining the various research methods considered, discussing their feasibility in addressing the research question. Next, we elaborate on the rationale behind selecting a mixed-methods approach and how it enabled a comprehensive examination of both usability outcomes and user satisfaction. Through the triangulation of data sources, we sought to provide a nuanced understanding of how predictability affects the interaction dynamics between humans and AI systems in immersive VR settings.

By detailing the experimental design, data collection tools, and analysis techniques, this section provides a thorough explanation of how the study was structured to yield meaningful insights into the role of predictability in AI-driven Virtual Reality.

3.1.1 Potential Research Methods

Several research methods were considered, each offering distinct advantages for answering your research question, including qualitative, quantitative, and mixed-methods approaches.

3.1.1.1 Qualitative Methods

Qualitative methods such as interviews, focus groups, or observations provide in-depth insights into users' subjective experiences and perceptions. This method could help understand how users perceive the AI's predictability in VR environments, their cognitive load, and their interaction dynamics with the system (Onwuegbuzie et al., 2009).

- Advantages: Qualitative data allows for exploring complex emotional, cognitive, and behavioral reactions to AI-driven systems, offering rich insights into user experiences.
- Challenges: These methods can be resource-intensive and time-consuming, especially in VR environments, where users' feedback may need to be captured in real time or immediately after VR exposure. Additionally, they often lack the statistical rigor required for generalization.

3.1.1.2 Quantitative Methods

Quantitative methods, such as surveys, task performance metrics, and behavioral observations, can be used to gather numerical data on variables such as user satisfaction, task completion time, and error rates. This method is crucial for testing hypotheses regarding the predictability and control users have over the AI system.

- Advantages: Quantitative research enables statistical analysis, offering objective data on user performance and experience. It allows for comparisons between different conditions, such as manual, predictable, and unpredictable AI-driven interactions in VR.
- Challenges: Quantitative methods alone might fail to capture the nuanced, subjective experiences of users. For instance, they may not explain why a user feels frustrated or satisfied with a certain interaction style, which qualitative methods can illuminate.

3.1.1.3 Mixed-Methods Approach

Given the complexity of our research question, a mixed-methods approach was considered most appropriate. Mixed-methods research combines the strengths of both qualitative and quantitative approaches to offer a comprehensive understanding.

- Advantages: This approach allows for the triangulation of findings. For instance, quantitative data such as task performance and error rates can be sup-

plemented by qualitative insights into why users struggled with certain tasks. Triangulation increases the validity of your findings by addressing research questions from multiple perspectives (R. Johnson & Onwuegbuzie, 2004).
- Challenges: Mixed-methods research can be resource-intensive, requiring careful planning to ensure the integration of qualitative and quantitative findings.

3.1.2 Feasibility of Potential Methods

When evaluating the feasibility of these methods, several factors were considered:
- Resource Constraints: Conducting extensive interviews or focus groups with each participant after a VR experiment would have required significant time and resources. Moreover, gathering qualitative data during VR interactions introduces challenges, as users may find it difficult to articulate their experiences while immersed in a virtual environment.
- Data Collection Tools: We considered platforms such as PsyToolkit and Prolific.com for data collection, which are well-suited to gathering large-scale quantitative data. For qualitative analysis, tools like Atlas.ti or NVivo were considered. However, Atlas.ti was preferred due to its team collaboration features and familiarity.
- Ethical Considerations: The potential for discomfort in VR environments was also factored in, making methods like interviews or real-time feedback potentially disruptive. This is why automated surveys post-interaction were chosen as a less intrusive method to collect qualitative feedback.

3.1.3 Chosen Research Methods

Ultimately, a mixed-methods approach was selected, combining quantitative analysis of learning outcomes and qualitative examination of user experiences. This decision was based on the need to:
- Gather objective performance metrics (e.g., task completion times, error rates, and SUS scores) through tools like R Studio for statistical analysis.
- Gain subjective insights into user experiences, which were analyzed using qualitative software like Atlas.ti.

The study design involved participants interacting with a VR prototype under three distinct conditions (manual, predictable, and unpredictable AI). Quantitative data, such as task performance and SUS scores, were supplemented by qualitative feedback from participants, offering insights into their experiences with each condition. This mixed-methods approach allowed for a comprehensive understanding of how

AI predictability influences user experience in VR.
By combining these methods, we could assess not only the task performance and usability of the system but also explore the subjective experiences that might not be captured through numbers alone. The triangulation of qualitative and quantitative data provided a more robust and nuanced analysis.

3.1.4 Rationale for Choosing Mixed-Methods Approach

The decision to use a mixed-methods approach was motivated by several factors:

- Complexity of the Research Question: The research question required understanding both performance metrics (e.g., error rates and task times) and deeper cognitive/affective responses to VR-based AI predictability. No single method could provide a complete answer.
- Enhancing Validity: By using both types of data, the research design aimed to ensure triangulation, where findings from one method (e.g., qualitative feedback) could validate findings from another (e.g., quantitative performance metrics).
- Time and Resource Efficiency: Given constraints, automated tools for collecting both types of data were used to streamline the process. Tools like PsyToolkit for pre-study surveys and Flutter for developing web apps to automate SUS questionnaires were efficient and scalable, supporting the collection of both quantitative and qualitative data with minimal researcher intervention.

The mixed-methods approach allowed for a multi-faceted exploration of AI predictability in VR, balancing the depth of qualitative insights with the rigor of quantitative data. This method was chosen to ensure that the research captured both the objective performance metrics and the subjective experiences of users, providing a comprehensive understanding of the research question.

3.2 System/Software

In this study, we utilized various software tools for prototyping, data collection, and analysis. Below, we outline the tools considered, their feasibility, and the rationale behind selecting the ones we used.

3.2.1 Prototyping

For our prototyping, we chose VR because it is an immersive technology in the gaming world, offering a new and challenging experience in our learning process.

Additionally, it allows for convenient control of environmental variables and the simulation of real-world experiences.

3.2.1.1 Virtual Reality

A virtual reality framework is a customizable application comprising design patterns and components that aid virtual reality developers through modularity, reusability, and extensibility (Protopsaltis & Papagiannakis, 2020). We aimed to pursue prototyping with Virtual Reality to control environmental variables. We were particularly intrigued by the work of (Weistroffer et al., 2013) in assessing the acceptability of Human-Robot Collaboration using Virtual Reality, with the potential for gaining further insights and exploring different scenarios from the user's perspective. However, in our case, we inclined towards predictability and AI.
We were quite excited about exploring VR for our project though our work could have been completed without using it too. As a user most of us are fascinated by the VR technology and now we had the opportunity to develop a prototype in VR ourselves. We were able to learn how much of an effort one has to put to develop an application in VR and what could go wrong any minute.
To construct the VR prototype, we had various tools at our disposal, including Unity, Unreal Engine, among others, which we have considered.

3.2.1.2 Unity

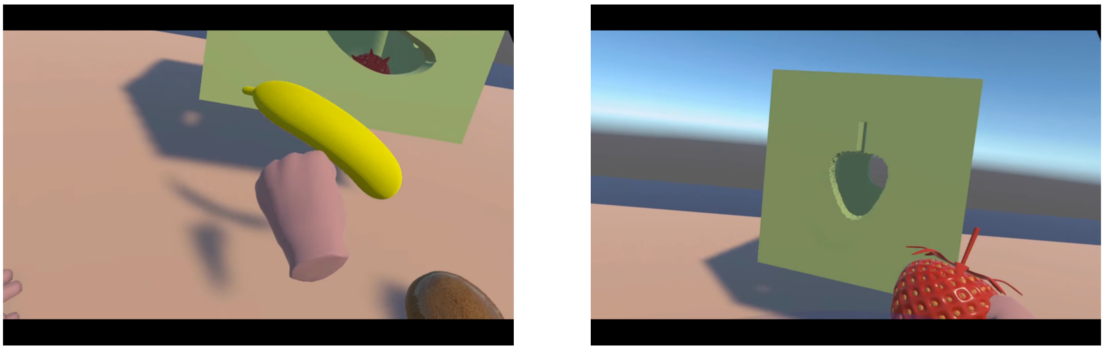
Unity is a versatile cross-platform game development engine introduced by Unity Technologies in 2005. It enables developers to create 2D, 3D, augmented reality (AR), and virtual reality (VR) applications, supporting a wide range of platforms such as PC, Mac, Android, iOS, WebGL, and more. Boasting a robust graphics engine, built-in physics simulation, and a C-Sharp scripting environment, Unity facilitates the creation of visually stunning and interactive experiences.

While other engines, like Unreal Engine, were considered, Unity was chosen for its ease of use, especially with C scripting, which aligned well with the team's technical expertise. Furthermore, Unity's strong support for VR applications and rapid prototyping made it the best option for our research goals.

3.2.1.3 Blender

Blender is a versatile, open-source 3D creation suite that stands out for its wide-ranging capabilities and cost-effectiveness.

We chose Blender because it offers a powerful, all-in-one solution for 3D creation, and it's completely free. Its wide range of tools like modeling, animation, texturing, rendering with Cycles and Eevee, and even video editing—makes it perfect for both beginners and professionals. Being open-source, Blender is constantly updated by its passionate community, providing access to cutting-edge features. We also appreciate its flexibility, with Python scripting allowing for customization and cross-platform compatibility, which suits various workflows. While other software like Maya or ZBrush are specialized for certain industries, Blender delivers professional-level capabilities without the steep cost or complexity. Below are some images in Figure 3.1 of prototype designed using blender and imported in unity.



Participant grabbing and inserting fruits

Participant observing while cube is rotating

Figure 3.1: Prototype designed using blender and unity

Compared to other options, Blender stood out because it provides everything We needed without the financial investment. While Maya excels in character animation, Cinema 4D is great for motion graphics, and ZBrush is the best for detailed sculpting, Blender covers all these areas well enough for most of our project. The constant innovation, like the real-time Eevee engine and 2D/3D hybrid Grease Pencil, made it the obvious choice for us.

3.2.2 Data Collection

Regarding the data collection we went through a lot of tools and techniques, comparing those to each other and finding best suitable ones for us. We did not go with the interview because we think giving participants their time to answer questions using the app would be more convenient and stress free.

3.2.2.1 PsyToolkit

PsyToolkit is a free to use toolkit for programming and running experiments and surveys. We used it to create a survey for our pre study as it is the only free website offering running programmable online psychological experiments and surveys also many students and academics around the world are using it. We had to implement a script that can be seen in the Figure 3.2 to create the survey.

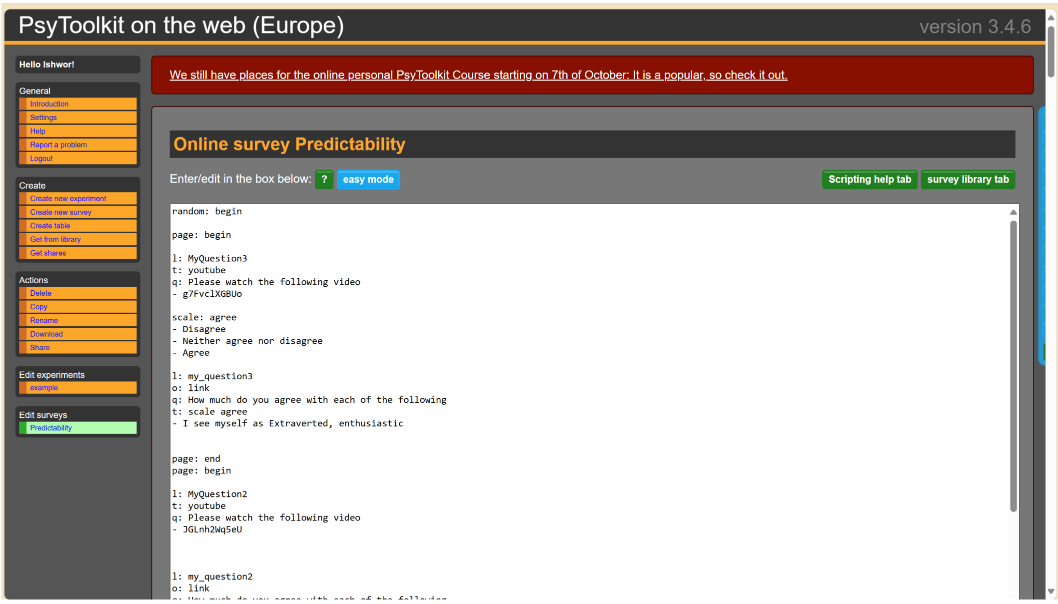


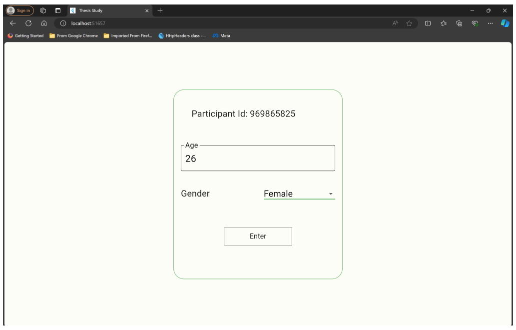
Figure 3.2: Designing the survey using PsyToolKit

3.2.2.2 Prolific.com

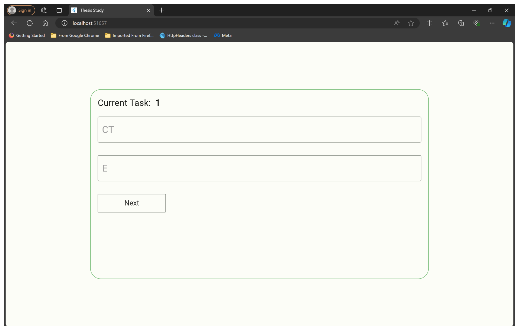
Prolific.com is a platform that connects researchers with participants for online studies. It is widely used by academic researchers, businesses, and organizations to conduct surveys, experiments, and market research. We did try other sites to recruit participants for online survey but those barely had one participant in a day which made us choose Prolific.com.

3.2.2.3 Flutter

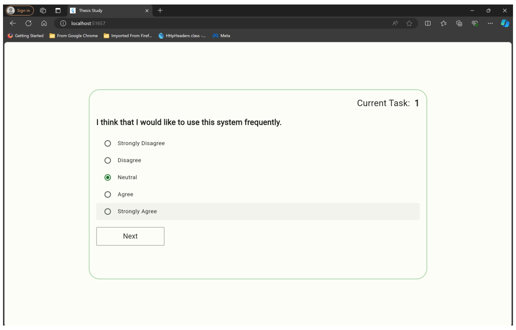
Flutter is a framework in dart programming language which helps to build a cross platform application with a single code base. We used it to build a web app which had the SUS Questionaires and one other question for the quantitative analysis. We decided to build the app so that the calculation of SUS score would be automated else we had to use other websites built in github which were not so transparent about how they calculated the score. Also, we used flutter to build the web app because we already had an experience of using flutter for developing simple app faster and conveniently in our previous course. Some images of the web app are shown in Figure 3.3.



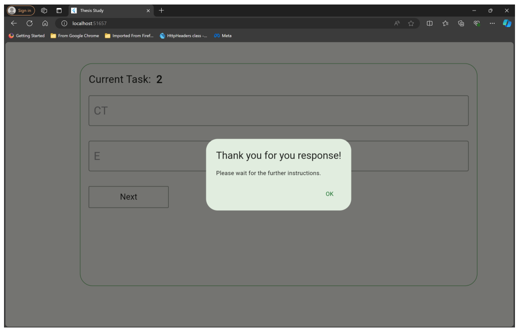
Page for the demographic data in the App



Page for the errors and time taken



Sus Question



Completion

Figure 3.3: Some images of the Web App taken during the experiment

3.2.3 Analysis and Data Visualization

After conducting the experiments we had quantitative and qualitative data on our hands. We needed some tools and softwares to analyze and visualize the data. For the quantitative analysis we had options like Python, R programming etc and for qualitative analysis we had options like NVivo, Atlas.ti etc.

3.2.3.1 R Studio

R Studio is an integrated development environment (IDE) designed for the R programming language, offering a comprehensive set of tools for statistical computing, data analysis, and visualization. With a user-friendly script editor, interactive console, and integrated help and documentation, R Studio facilitates a seamless workflow for users ranging from beginners to experienced statisticians.
While alternatives like Python were considered, R Studio was chosen for the quantitative data analysis and visualization as shown in the Figure 3.4, because of the

team's familiarity with R programming and its advanced statistical libraries. It provided the necessary tools to perform ANOVA and other statistical tests efficiently.

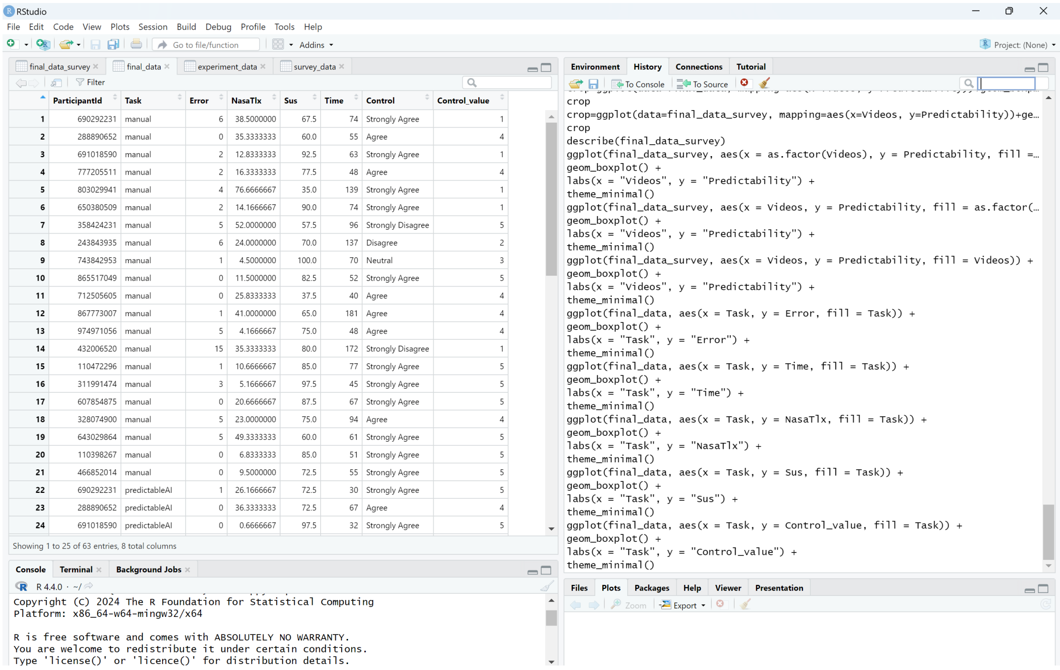


Figure 3.4: Analyzing the result from experiment using R programming

3.2.3.2 Atlas.ti

Atlas.ti is a qualitative data analysis (QDA) software tool widely employed in social sciences and research fields. It enables researchers to code, categorize, and analyze diverse qualitative data, including text, images, audio, and video. The software supports complex queries, allowing users to extract meaningful patterns and insights from large datasets, and facilitates network and relationship analysis to uncover connections within the qualitative data. Example map of the code generated using AI has been shown in Figure 3.5.

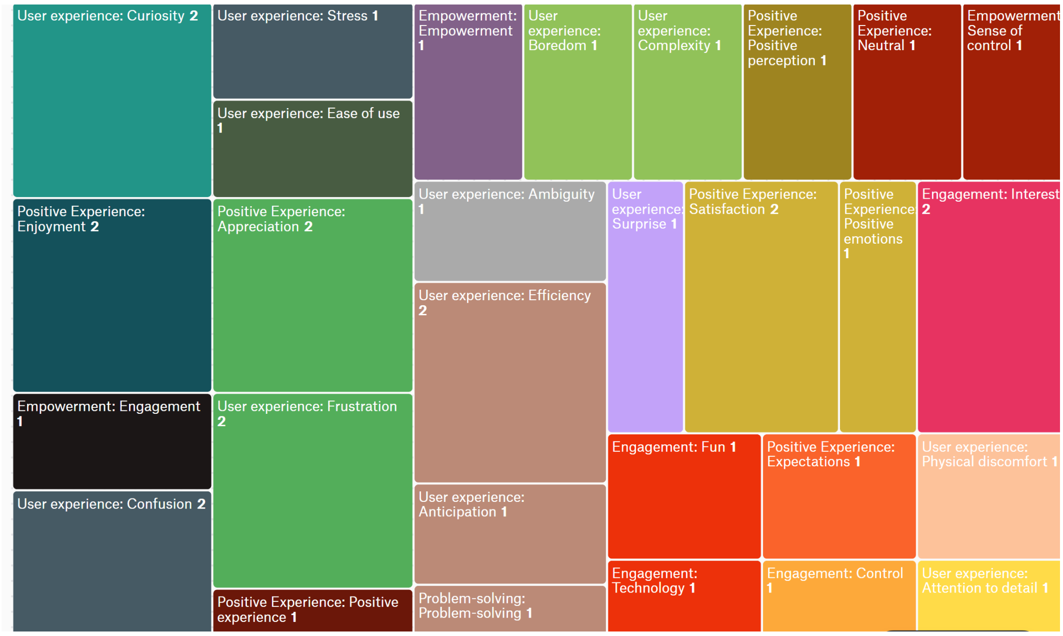


Figure 3.5: Qualitative analysis map using Atlas.ti

Though other options like NVivo offered both qualitative and quantitative analysis, atlas.ti offers quering and team collaboration which we much prefered since we have R studio for the quantitaive analysis. Also the past experience with Atlas.ti was helpful in this study.

3.2.4 Roles and Responsibilities in Thesis Development

Table 3.1: Thesis Work Contribution

Task	Details of Contribution	Contributor
Thesis Planning, Literature Review	Discussion on the project formulating a Research Question between us and our supervisor, Diving into the various articles to establish background and theory for the project and finding some relevant preceding research	Ishwor & Ismail
Study Design Documentation	Designing the study including our research questions, research methods, and tools we decided to use.	Ismail
Prototype Design	Design was created using Blender and then imported in Unity	Ismail
Algorithm Implementation for the Prototype	Using C-Sharp Scripting rotation algorithm was implemented	Ishwor
Creating Survey for pre study, Data Analysis of the Survey	Recorded three videos of the prototype testing in three conditions and created a survey using Psytoolkit and published in Prolific.com, Analysis of the data using R programming in Rstudio using ANOVA, ART Anova tests.	Ishwor & Ismail
Web app for Main Study Data Collection	Made a web based application for the SUS Questionaires and Control related question which can calculate the score and convert data in csv form	Ishwor
Conducting Pilot Study & Main Study	Before running a main study we had 2 participants for the pilot study who helped us to refine main study with possible errors and suggestions. Main Study was conducted following the pilot study with 21 participants.	Ishwor & Ismail
Data Analysis of the Main Study, Defense, Report Finalization	Collected data were analyzed with Rstudio and Atlas.ti. Final report drafting and work around was done after the defense.	Ishwor & Ismail

4

Method

4.1 Pre-Study

The pre-study aimed to operationalize the concept of "predictability" within the context of AI-driven virtual reality (VR) environments. To achieve this, we employed Rule-based AI, also known as Symbolic AI, to control the behavior of virtual objects within the VR setting. Rule-based AI functions through predefined "if-then" logic, allowing us to create clear and distinct conditions for the study. In the context of this pre-study, Rule-based AI was crucial because it enabled us to design predictable and unpredictable scenarios by setting fixed rules that governed system behavior.

For the predictable condition, the AI followed a specific rule: upon interaction with a virtual object, the system would rotate a cube along the z-axis in a consistent, expected manner. In contrast, the unpredictable condition involved the Rule-based AI applying different rules, where the cube would rotate randomly across multiple axes.

The use of Rule-based AI was essential in this pre-study because it provided a controlled, repeatable method for manipulating the predictability of the VR environment. By relying on a fixed set of rules, we ensured that the experimental conditions were systematically applied across all participants, allowing us to reliably measure and compare perceptions of predictability. The findings from the pre-study showed that participants found the predictable condition significantly more predictable, as expected, confirming that the Rule-based AI successfully created the desired distinction between the two conditions.

This pre-study was a crucial step in establishing a foundation for the main study, as it validated our approach to manipulating predictability using Rule-based AI, providing a clear framework for assessing its impact on user experience.

4.1.1 Study Design

The pre-study employed a within-subjects design, where each participant evaluated the predictability of three distinct conditions of cube rotation in a VR environment. The conditions were: manual rotation, predictable rotation along the z-axis, and unpredictable rotation along all three axes. This design allowed for direct comparison of participants' perceptions of predictability across the different conditions, ensuring that each participant experienced all variations and provided comparative evaluations.

4.1.2 Participants

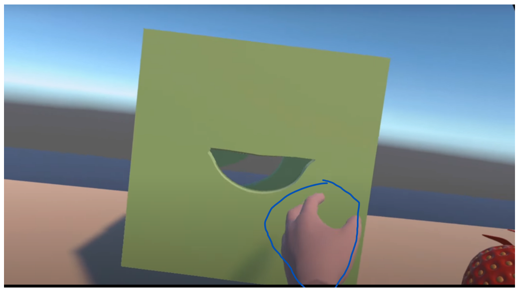
A total of 26 participants (15 females and 12 males) were recruited through Prolific, compensated £1.05 each, and directed to PsyToolkit to evaluate the predictability of different conditions through a survey.

4.1.3 Materials and Instruments

The pre-study utilized PsyToolkit, a comprehensive software toolkit designed for programming and running cognitive-psychological experiments and surveys. PsyToolkit was chosen for its robustness and flexibility, allowing for precise control over the experimental procedures and efficient data collection. Participants evaluated the predictability of cube rotations presented through video demonstrations of three conditions in a VR prototype:

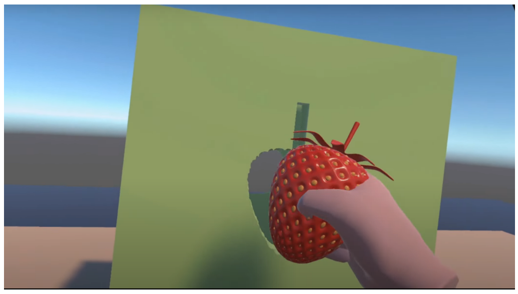
- **Manual Rotation:** The cube remains stationary without any rotational movement upon picking up a fruit. The cube was then moved manually by one of the research team members that can be seen in Figure 4.1.

1. Participant starts picking the fruit



2. In the next step, participant grabbing the cube

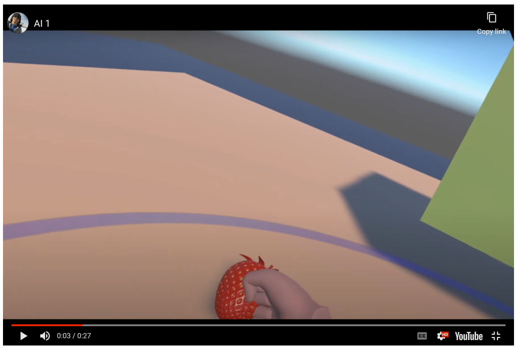
3. Participant manually rotating the cube for right side



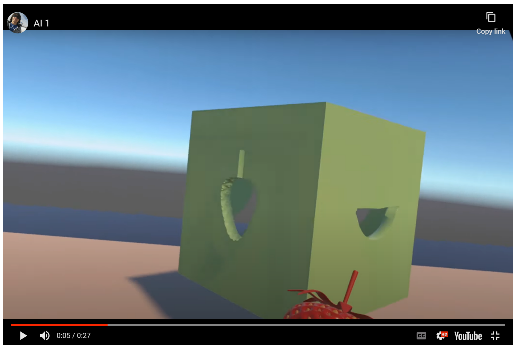
4. Participant inserting the fruit in the right shaped hole

Figure 4.1: 2x2 Grid of participant performing under manual condition

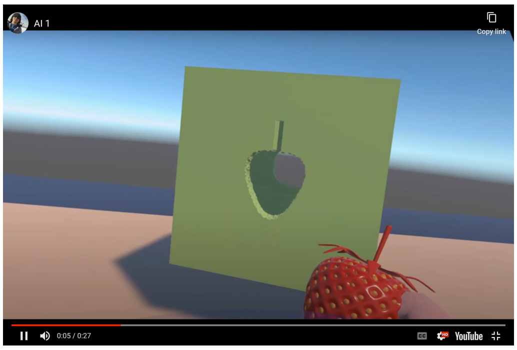
- **Z-Axis Rotation (Predictable):** The cube rotates automatically only along the z-axis (vertical axis) upon picking up a fruit and stops with the corresponding face to the picked-up fruit which can be seen in Figure 4.2.



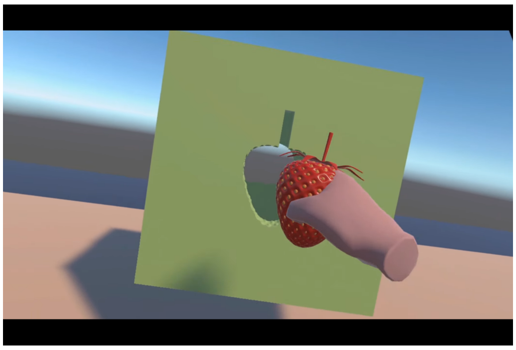
1. First, participant picking the fruit



2. Cube rotating on Z-axis only to show the right face



3. Cube almost completing the rotation to show right face

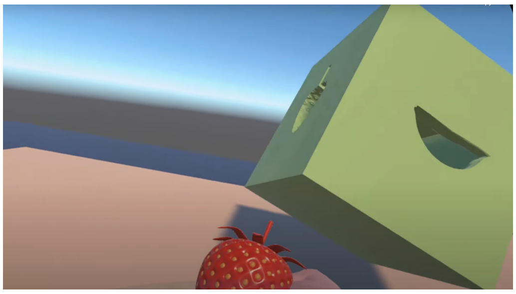


4. Participant inserting fruit in the correct fruit shaped hole

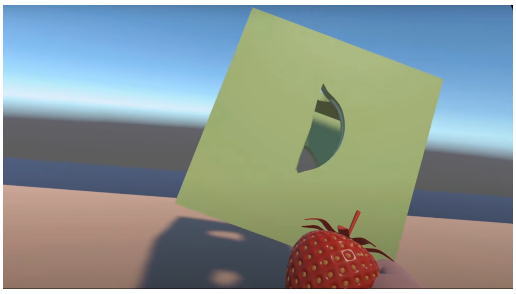
Figure 4.2: 2x2 Grid of participant performing under predictable condition

- **Random Rotation (Unpredictable):** The cube rotates automatically along all three axes (x, y, z) upon picking up a fruit and then stops with the corresponding face to the picked-up fruit that can be seen in Figure 4.3.

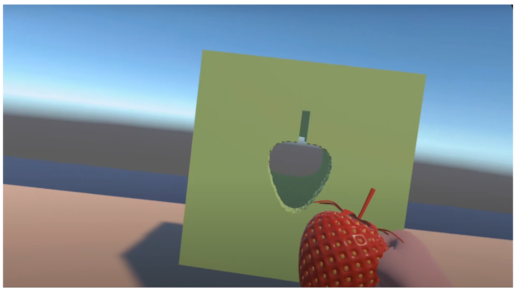
1. First, participant picking the fruit



2. Cube rotating in all axes randomly



3. Cube continues to rotate in all axes randomly until it show the correct hole



4. Participant inserting the fruit after correct shaped hole is shown

Figure 4.3: 2x2 Grid of participant performing under unpredictable condition

4.1.4 Procedure

Participants were directed to the PsyToolkit platform, where they watched video clips demonstrating the three conditions of cube rotation. Following each video, participants were asked to rank the predictability of the cube's movement on a scale from 1 to 7. The procedure was structured to ensure clarity and ease of completion. Participants began by reading an introduction that briefed them on the study's purpose and provided instructions on how to complete the survey. They then viewed three video clips, each depicting one of the rotation conditions. After watching each video, participants rated the predictability of the cube's movements on a scale from 1 to 7. Their responses were automatically recorded and stored for analysis. The entire process was designed to be user-friendly and efficient, ensuring that participants could easily understand and complete the tasks.

4.1.5 Data Analysis

Data were collected through PsyToolkit into an Excel file and analyzed using R Studio. Descriptive statistics (mean and standard deviation) and inferential statistics (ART-ANOVA and Tukey's Honest Significant Difference tests) were used to determine significant differences between conditions.

4.1.6 Results

Preliminary results indicated varying levels of predictability among the conditions, which helped refine the study design and ensure clarity for participants in the main study.

4.1.6.1 Descriptive Statistics

Descriptive statistics showed that the mean predictability ratings for the conditions were as follows: Manual Condition ($M = 5.04, SD = 1.99$), Predictable Condition ($M = 6.31, SD = 1.19$), and Unpredictable Condition ($M = 4.69, SD = 2.13$).

Table 4.1: Descriptive Data from Survey

Variable	N	Mean	SD	Median	Trimmed	MAD	Min	Max
id	26	13.00	7.21	13.00	13.00	8.90	1.00	25.00
gender	26	1.58	0.50	2.00	1.59	0.00	1.00	2.00
age	26	31.46	10.60	28.00	29.77	5.93	22.00	68.00
manual	26	5.04	1.99	6.00	5.23	1.48	1.00	7.00
predictableAI	26	6.31	1.19	7.00	6.55	0.00	3.00	7.00
unpredictableAI	26	4.69	2.13	5.00	4.77	2.97	1.00	7.00

4.1.7 Quantitative Results

The ART ANOVA results indicated a significant main effect of rotation condition on predictability ratings, $F(2, 75) = 5.45$, $p = 0.006$. Tukey's HSD post-hoc tests revealed the following:

- The predictable condition was rated significantly higher in predictability compared to the manual condition (mean difference = 1.27, 95 % $CI[0.06, 2.48]$, $p = 0.037$).
- There was no significant difference in predictability ratings between the unpredictable condition and the manual condition (mean difference = -0.35, 95 % CI[-1.55, 0.86], $p = 0.772$)
- The unpredictable condition was rated significantly lower in predictability compared to the predictable condition (mean difference = -1.62 , 95 % CI[-2.82, -0.41] , $p = 0.006$).

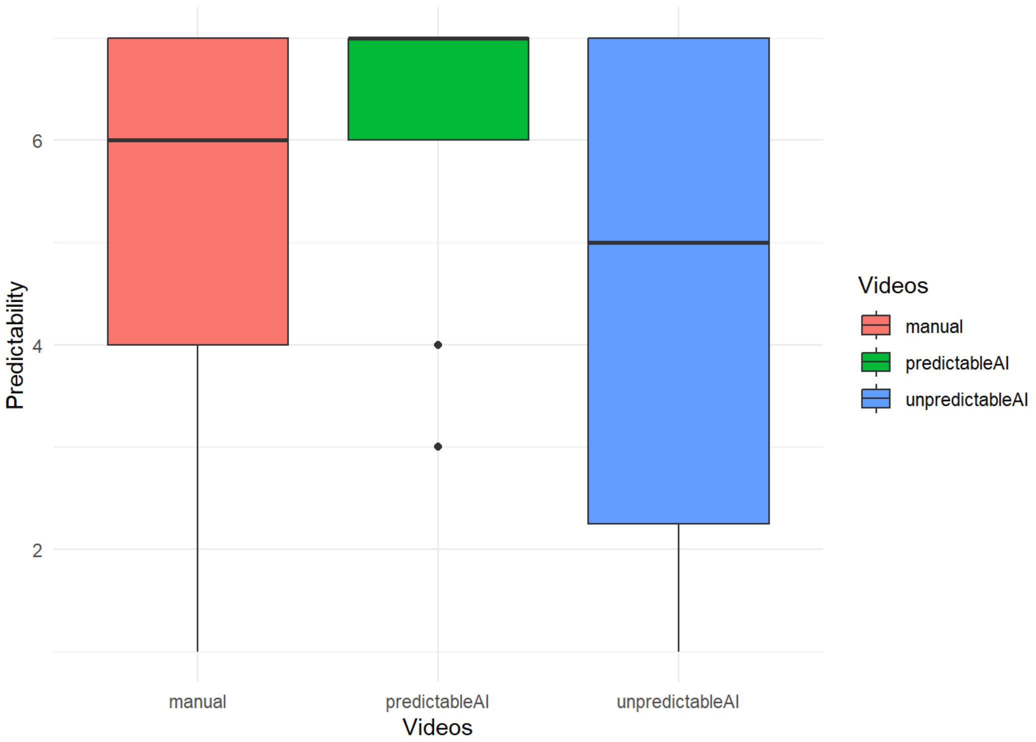


Figure 4.4: Box Plot in R of Predictability under three conditions

4.1.8 Discussion

The pre-study successfully operationalized the concept of predictability by providing clear, measurable conditions for participants to evaluate. This success is evidenced by the significant differences in predictability ratings between the conditions. The distinct rotation patterns for each condition were easily distinguishable, allowing participants to make clear judgments about predictability. This clarity was crucial in ensuring that the main study could build on a solid foundation where the concept of predictability was well-understood and reliably measured.

The predictable rotation condition, limited to the z-axis, received high predictability ratings due to its consistent and straightforward nature. In this condition, the cube's behavior was regular and repeatable, making it easy for participants to anticipate the outcome after picking up a fruit. This aligns with cognitive theories that emphasize the importance of consistency in enhancing predictability (Bubic et al., 2010). When users can easily form expectations about system behavior, their cognitive load is reduced, and their interactions become more efficient and satisfying.

Conversely, the unpredictable rotation condition received lower predictability ratings. This condition involved the cube rotating along all three axes in a seemingly random manner, which introduced variability and complexity that participants

found difficult to predict. The lack of a clear pattern in the cube's movement likely led to frustration and confusion among participants, highlighting the importance of predictability in system design.
The manual condition also received lower predictability ratings, which can be attributed to the variability introduced by human control. Unlike the automated conditions, the manual rotation depended on the actions of the research team member, which could introduce inconsistencies. D. Norman (2014) discusses how user expectations are crucial in determining predictability; when the system's behavior varies due to human intervention, it may be perceived as less predictable. Additionally, manual control requires users to predict their actions and the system's responses, which can increase cognitive load compared to automated, predictable behavior (Lee & See, 2004).
These findings underscore the importance of designing AI systems that are both consistent and predictable to enhance user satisfaction and performance. The pre-study's methodology and results provide a solid foundation for understanding how predictability influences user experience, guiding future design and implementation strategies for more intuitive and efficient AI-driven environments.

4.2 Pilot-Study

The pilot study aimed to test and refine the study procedures, software, and instruments used in the main study. This phase was essential for identifying and addressing potential issues that could impact the reliability and validity of the main study's results. By conducting a pilot study, we aimed to ensure that all elements of the experimental setup functioned as intended and that participants could easily understand and complete the tasks.

4.2.1 Study Design

The pilot study employed a within-subjects design, where participants experienced and evaluated three distinct conditions of cube rotation in a VR environment: manual rotation, predictable rotation along the z-axis, and unpredictable rotation along all three axes. This design allowed for a thorough examination of the experimental procedures and provided an opportunity to refine the setup based on participant feedback and observations.

4.2.2 Participants

Two male participants with prior VR experience were recruited for the pilot study. The small sample size was sufficient for identifying technical issues and procedural problems without the need for extensive statistical analysis. Participants' familiarity with VR technology ensured that their feedback would be focused on the experimental setup rather than on basic VR acclimatization.

4.2.3 Materials and Instruments

The pilot study utilized a VR environment created using Unity and Blender 3D. The tasks involved cube and fruit interactions under three conditions of rotation. The Meta Quest Pro VR headset was used to provide an immersive experience. This setup was chosen to replicate the conditions of the main study ensuring that any identified issues would be relevant to the larger experimental context.

4.2.4 Procedure

Participants first read and signed a consent form and then familiarized themselves with the VR environment through a dummy phase. They experienced the three conditions of cube rotation (manual, predictable, and unpredictable) in a repeated-measures design. Each participant was exposed to all three conditions, with the order randomized to control potential order effects.

During the manual rotation condition, participants moved the cube themselves, experiencing full control over its movements. In the predictable rotation condition, the cube rotated automatically along the z-axis when a fruit was picked up, stopping with the corresponding face to the picked-up fruit. In the unpredictable rotation condition, the cube rotated automatically along all three axes, stopping with the corresponding face to the picked-up fruit.

4.2.5 Data Analysis

Observational data were collected to identify and address potential issues with the experimental setup. These observations included participants' interactions with the VR environment, any technical difficulties encountered, and participants' feedback on the clarity and usability of the instructions and tasks.

The first trial of the pilot study revealed several issues, including problems with the recording system, Unity bugs, and the need for clearer instructions. Based on these findings, the research team made necessary adjustments to the software and procedures. A second trial was conducted to verify that all identified issues had been resolved, confirming the feasibility and reliability of the study procedures, instruments, and software.

4.2.6 Outcomes and Modifications

The pilot study confirmed the feasibility and reliability of the study procedures, instruments, and software, ensuring that the main study could proceed smoothly.

4.2.7 Discussion

Conducting a pilot study is crucial for identifying and rectifying potential issues in the experimental setup, ensuring the main study's reliability and validity. The pilot study allows researchers to test the feasibility of their design, refine procedures, and gather preliminary feedback to improve the clarity and functionality of the tasks. This phase is particularly important in complex studies involving new technologies, such as VR, where unexpected technical and procedural challenges can arise.
In this study, the pilot phase was instrumental in highlighting and resolving issues related to software bugs, recording systems, and participant instructions. Identifying the errors like grabbing the cube while it's still rotating, throwing the fruits off the table, inserting all the fruits from a single hole, trying to keep the fruit on edge instead of inserting it all the way in, not inserting the fruits in correct order. By addressing these problems before the main study, the research team ensured a smoother, more efficient data collection process and enhanced the overall validity of the research. The pilot study's feedback and observations provided a solid foundation for the main study, contributing to its methodological soundness and the reliability of its outcomes.

4.3 Main Study

The main study aimed to investigate whether AI predictability significantly affects usability and overall workload in VR environments. Building on the findings from the pre-study and the pilot study, this phase sought to provide comprehensive insights into how different levels of AI predictability impact user experience and performance in a VR setting.

4.3.1 Study Design

The main study employed a within-subjects design, where participants experienced and evaluated three distinct conditions of cube rotation in a VR environment: manual rotation, predictable rotation along the z-axis, and unpredictable rotation along all three axes. This design allowed for direct comparison of participants' perceptions and performance across different conditions, ensuring robust and reliable results.

4.3.2 Participants

The primary study included 21 participants (13 females and 8 males) between the ages of 22 and 55 years (M = 30.29, SD = 10.31), with varying levels of VR experience. Participants were recruited via various social media platforms and were rewarded with mini French waffles for their participation. Inclusion criteria required participants to be fluent in English. There were no specific exclusion criteria. All participants provided informed consent prior to participation, and they were rewarded with mini French waffles for their participation. The diverse demographic of the participants provided a wide range of perspectives, enhancing the generalizability of the findings.

4.3.3 Materials and Instruments

The VR environment was designed using Unity and Blender 3D, with tasks involving cube and fruit interactions under three conditions of rotation. The Meta Quest Pro VR headset was used to provide an immersive experience.

4.3.4 Procedure

Participants first read and signed a consent form and then familiarized themselves with the VR environment through a dummy phase. They experienced the three conditions of cube rotation:

- **Manual Rotation:** The cube remains stationary without any rotational movement upon picking up a fruit. The cube was then moved manually by one of the participant.
- **Z-Axis Rotation (Predictable):** The cube rotates automatically only along the z-axis (vertical axis) upon picking up a fruit and stops with the corresponding face to the picked-up fruit.
- **Random Rotation (Unpredictable):** The cube rotates automatically along all three axes (x, y, z) upon picking up a fruit and then stops with the corre-

sponding face to the picked-up fruit.

The order of conditions was randomized for each participant to control for order effects.

In the manual rotation condition, participants moved the cube themselves, experiencing full control over its movements. In the predictable rotation condition, the cube rotated automatically along the z-axis when a fruit was picked up, stopping with the corresponding face to the picked-up fruit. In the unpredictable rotation condition, the cube rotated automatically along all three axes, stopping with the corresponding face to the picked-up fruit.

After completing each condition, participants filled out a set of questionnaires, including the System Usability Scale (SUS), a control question, the NASA Task Load Index (NASA-TLX), and an open-ended feedback question. The research team recorded the time taken, performance, and errors for each condition.

4.3.5 Data Analysis

Quantitative data were collected through the SUS, control question, NASA-TLX, and performance metrics. Qualitative data were gathered from open-ended feedback responses. The data were analyzed using R Studio for quantitative insights and Atlas.ti for qualitative themes. Descriptive statistics (mean and standard deviation) and inferential statistics (ANOVA and Tukey's Honest Significant Difference tests) were used to compare outcomes across different conditions.

4.3.6 Outcomes and Modifications

Data were analyzed using R Studio for quantitative insights and Atlas.ti for qualitative themes. Descriptive statistics (mean and standard deviation) and inferential statistics (ANOVA and Tukey's Honest Significant Difference tests) were used to compare outcomes across different conditions.

4.3.7 Results

This section presents the findings from the study, organized into descriptive statistics, quantitative results, and qualitative results. The descriptive statistics provide an overview of the key variables across different conditions, illustrating the central tendencies and variability within the data. The quantitative results explore the statistical significance and relationships between variables using appropriate statistical tests. Finally, the qualitative results offer insights into participants' experiences and perceptions, enhancing the understanding of the quantitative findings.

4.3.7.1 Descriptive Statistics

Descriptive statistics for the main study showed the following means and standard deviations for the three conditions.
For the Manual condition, the mean of being in control score was $M = 3.52, SD = 1.63$; the mean duration to complete the task was $M = 80.90$ seconds, $SD = 41.58$; the mean number of errors was $M = 3.00, SD = 3.52$; the mean NASA Task Load Index (NASA-TLX) score was $M = 24.63, SD = 18.99$; and the mean System Usability Scale (SUS) score was $M = 73.93, SD = 17.37$.
For the Predictable condition, the mean of being in control score was $M = 4.05, SD = 1.47$ the mean duration to complete the task was $M = 44.48, SD = 20.09$; the mean number of errors was $M = 2.24, SD = 2.74$; the mean NASA Task Load Index (NASA-TLX) score was $M = 18.76, SD = 14.33$; and the mean System Usability Scale (SUS) score was $M = 78.45, SD = 13.54$.
For the Unpredictable condition, the mean of being in control score was $M = 3.33, SD = 1.65$; the mean duration to complete the task was $M = 65.67$ seconds, $SD = 30.21$; the mean number of errors was $M = 3.71, SD = 5.59$; the mean NASA Task Load Index (NASA-TLX) score score was $M = 24.43, SD = 21.87$; and the mean System Usability Scale (SUS) score was $M = 66.43, SD = 22.07$.

Table 4.2: Descriptive Data from Experiment

Variable	N	Mean	SD	Median
Time_Manual	21	80.90	41.58	67.00
Errors_Manual	21	3.00	3.52	2.00
SUS_Manual	21	73.93	17.37	75.00
Control_Manual	21	3.10	1.48	4.00
Tlx_Manual	21	24.63	18.99	20.67
Time_Predictable	21	44.48	20.10	38.00
Errors_Predictable	21	2.24	2.74	1.00
SUS_Predictable	21	78.45	13.54	80.00
Control_Predictable	21	2.67	0.86	3.00
Tlx_Predictable	21	18.76	14.33	19.00
Time_Unpredictable	21	65.67	30.21	59.00
Errors_Unpredictable	21	3.71	5.59	1.00
SUS_Unpredictable	21	66.43	22.10	75.00
Control_Unpredictable	21	2.57	1.03	3.00
Tlx_Unpredictable	21	24.43	21.90	20.17
Control_value_Manual	21	3.52	1.63	4.00
Control_value_Predictable	21	4.05	1.47	5.00
Control_value_Unpredictable	21	3.33	1.65	4.00

4.3.7.2 Quantitative Results

4.3.7.2.1 Control The repeated-measures ANOVA for the control variable showed no significant main effect of task condition, $F(2, 40) = 1.53$, $p = 0.228$. Mauchly's test indicated a violation of sphericity ($p = 0.012$), but Greenhouse-Geisser ($\epsilon = 0.730$, $p = 0.232$) and Huynh-Feldt ($\epsilon = 0.773$, $p = 0.232$) corrections did not alter the non-significant result.

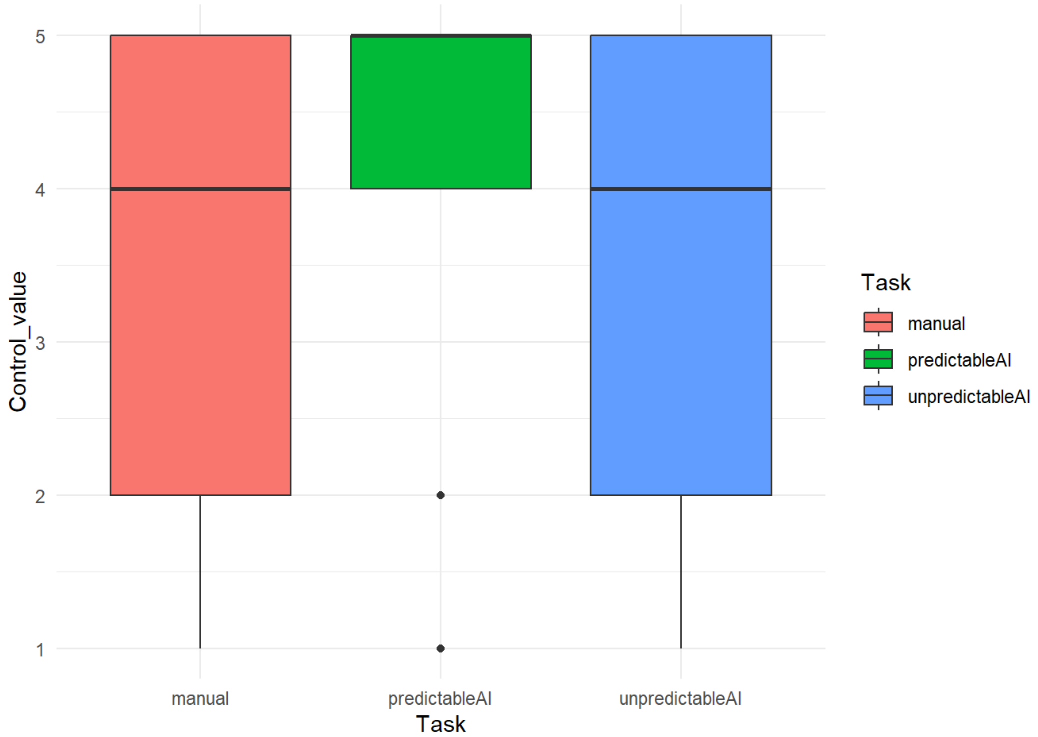


Figure 4.5: Box Plot in R of Control under three conditions

4.3.7.2.2 Duration The repeated-measures ANOVA for task completion time revealed a significant main effect of task condition, $F(2, 40) = 9.00$, $p < 0.001$. Mauchly's test indicated that the assumption of sphericity was violated ($p = 0.018$), so Greenhouse-Geisser ($\epsilon = 0.744$, $p = 0.002$) and Huynh-Feldt ($\epsilon = 0.790$, $p = 0.002$) corrections were applied, confirming the significant effect. Tukey's HSD post-hoc tests indicated the following pairwise comparisons for task completion time:

- Manual vs. Predictable: mean difference = 36.4, $SE = 8.48, t(20) = 4.30$, $p = 0.001$
- Manual vs. Unpredictable: mean difference = 15.2, $SE = 10.64, t(20) = 1.43$, $p = 0.344$
- Predictable vs. Unpredictable: mean difference = -21.2, $SE = 6.16, t(20) = -3.44$, $p = 0.007$

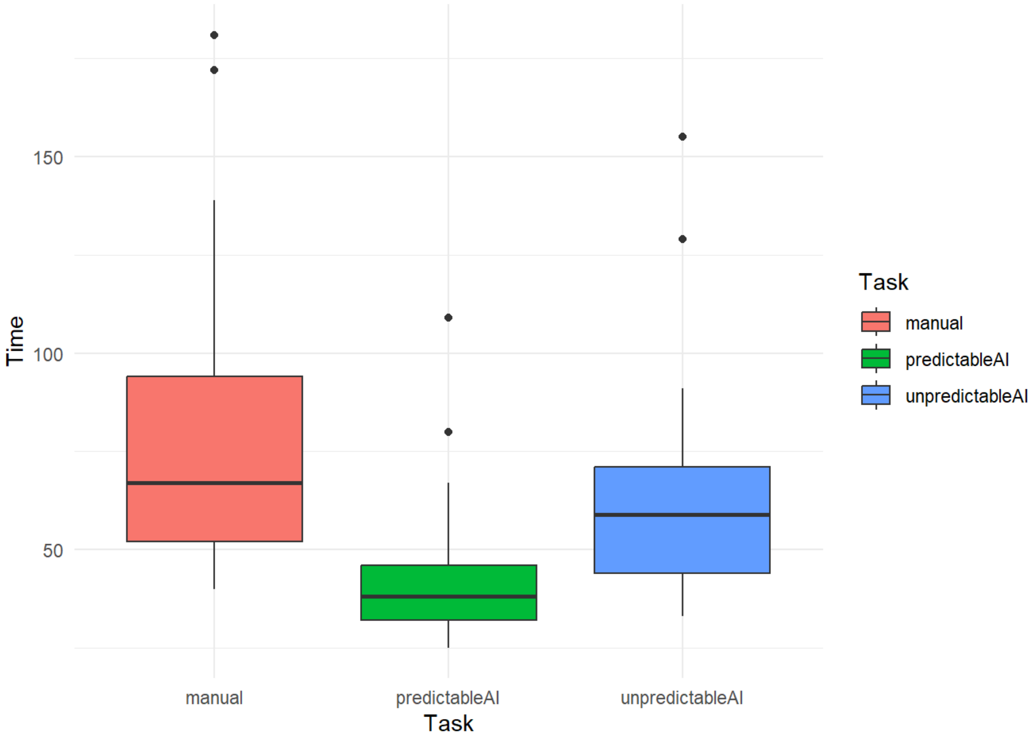


Figure 4.6: Box Plot in R of Time Duration under three conditions

4.3.7.2.3 Error Rate The repeated-measures ANOVA for error rates indicated no significant main effect of task condition on errors, $F(2, 40) = 1.09$, $p = 0.346$. Mauchly's test indicated that the assumption of sphericity had been violated ($p = 0.0029$), therefore, Greenhouse-Geisser ($\epsilon = 0.686$, $p = 0.328$) and Huynh-Feldt ($\epsilon = 0.719$, $p = 0.331$) corrections were applied, but the results remained non-significant.

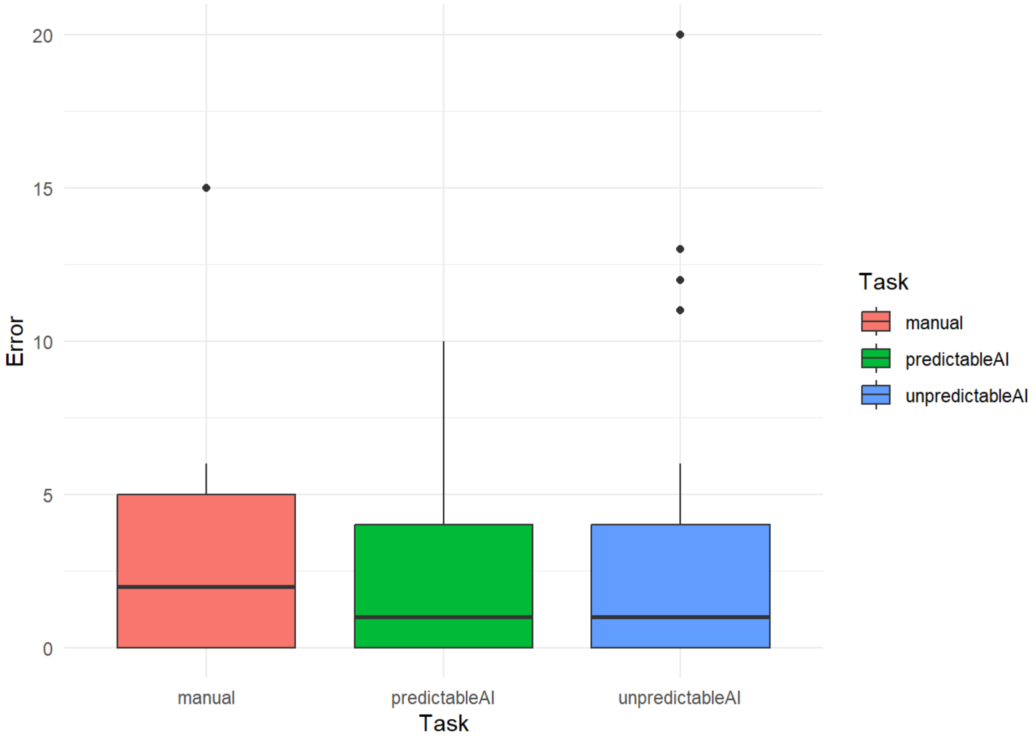


Figure 4.7: Box Plot in R of Error Rate under three conditions

4.3.7.2.4 NASA Task Load Index The repeated-measures ANOVA for NASA-TLX scores indicated a marginally significant effect of task condition, $F(2, 40) = 2.84$, $p = 0.070$. Mauchly's test suggested that the sphericity assumption was not violated ($p = 0.077$). However, Greenhouse-Geisser ($\epsilon = 0.809$, $p = 0.083$) and Huynh-Feldt ($\epsilon = 0.869$, $p = 0.079$) corrections were applied.

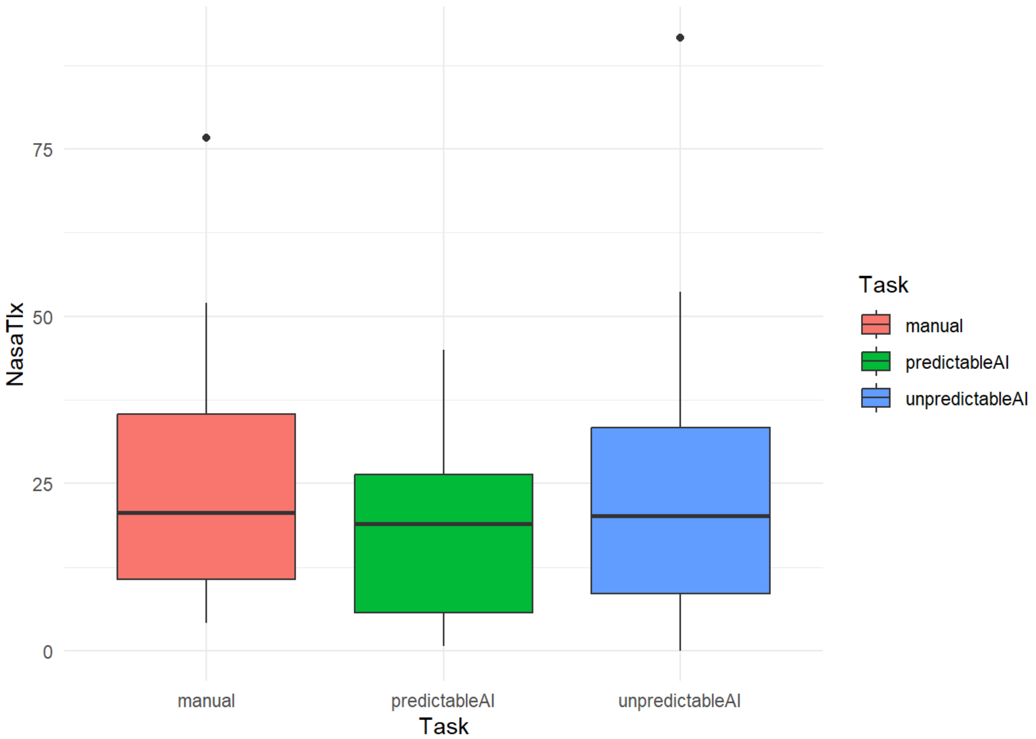


Figure 4.8: Box Plot in R of NASA-TLX under three conditions

4.3.7.2.5 SUS The ART ANOVA for SUS scores showed no significant main effect of task condition on usability ratings, $F(2, 60) = 1.47$, $p = 0.237$.

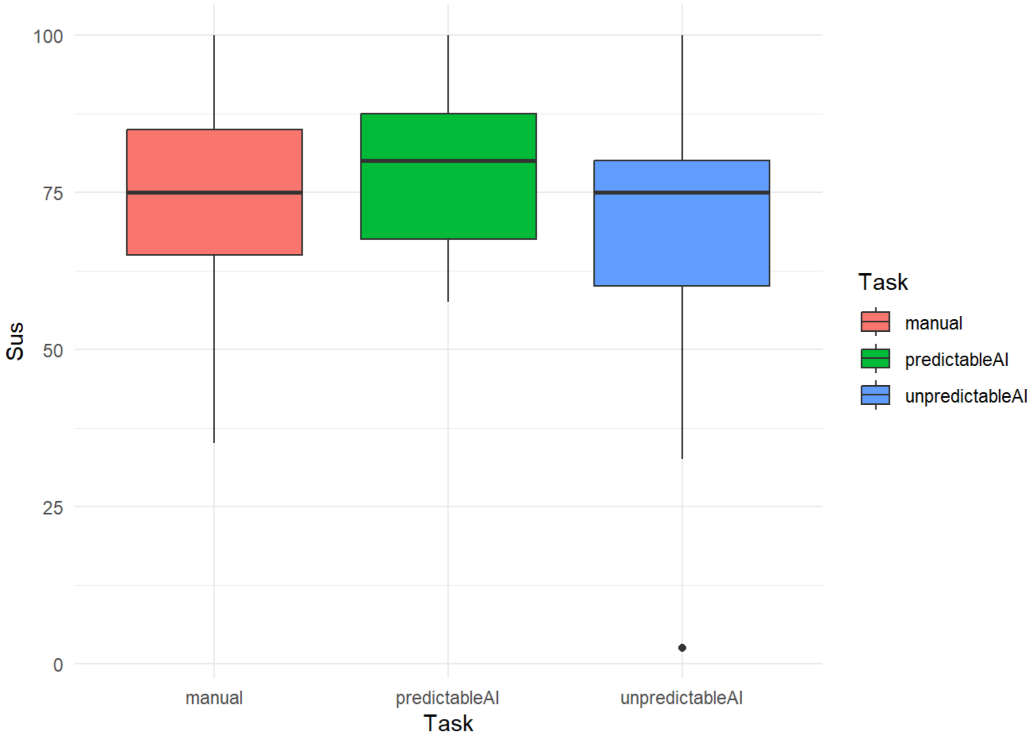


Figure 4.9: Box Plot in R of SUS under three conditions

4.3.7.3 Qualitative Results

As mentioned in the Method section, we had one open-ended question, to which 13 participants responded for each condition. The reason we only received responses from 13 participants, despite there being 21, is that we developed an application ourselves to store the responses on the cloud Firestore (Chougale et al., 2022). However, due to a minor bug in the app, which we discovered later, the responses of 8 other participants were not stored. We promptly fixed the issue upon noticing it and were able to collect responses from the remaining 13 participants.

4.3.7.3.1 Manual The participants in the manual condition exhibited a spectrum of reactions, encompassing expectations, surprises, and frustrations. While some embraced the sense of agency and empowerment derived from physically interacting with the VR environment, others encountered unexpected challenges, resulting in feelings of frustration and ennui. Overall, participants stressed the importance of clear task instructions and expressed a desire for increased novelty to sustain engagement throughout the trials.

Within this condition, participants articulated a variety of sentiments regarding their encounters. One participant observed, "**The task felt more intricate compared to the preceding ones. Although I experienced visual discomfort, I found the task compelling.**" Another participant reflected, "**Initially taken aback by the need to manipulate objects manually, I quickly adapted and found myself more adept and in control.**"
Furthermore, a participant shared, "**Engaging in manual interactions left me feeling invigorated and eager to explore subsequent tasks. It added an exciting dimension to the experience.**" Conversely, another participant voiced frustration, commenting, "**The task failed to evoke any significant emotions. It felt mundane, reminiscent of childhood activities. While the VR aspect added novelty, encountering glitches such as items falling through solid surfaces was irksome.**"
These testimonials underscore the diverse array of experiences and emotions participants encountered during the manual condition of the study, highlighting the need for nuanced considerations in future VR interactions.

4.3.7.3.2 Predictable AI Participants in the predictable AI condition generally found the task to be repetitive, yet they appreciated its simplicity and clarity. While some reported increased task proficiency and confidence through repetition, others expressed frustration due to delays in rotations and occasional errors. Despite these challenges, participants acknowledged the potential of the VR system for research and education, recognizing both its promising features and areas needing improvement in terms of clarity and engagement.
Within the Predictable AI condition, participants exhibited a spectrum of reactions to their experiences. Some expressed contentment with the predictability of the AI system, noting, "**It was still nice, not complex, and feels like it could be used to do some very meaningful tasks in the future.**" Another participant found solace in the tranquility of the task, remarking, "**The moment I started to experience the task, it somehow felt like I was living in my own little world. Everything felt so calm, and it was just me in control of my surroundings.**"
However, not all participants shared these sentiments. One participant emphasized the monotony of the task, commenting, "**I felt like I was doing the same task as in the first one except for the pattern of the rotation of the cube. I felt it was a repetitive task.**" Another participant expressed initial hesitation followed by growing confidence, stating, "**I think I felt confident about what**

to do, even though before starting the task I was unsure. Things were clear when the task started, and as I went on with the task, I gained confidence in my knowledge of what to do next."
These quotes encapsulate the varied responses participants had to the Predictable AI condition, portraying both positive and negative facets of their experiences.

4.3.7.3.3 Unpredictable AI In the unpredictable AI condition, participants expressed a mix of curiosity and interest in exploring the capabilities of the VR equipment. Some found automatic rotations and unpredictable behaviors engrossing, feeling a sense of control and focus. However, others encountered frustration and stress, particularly when faced with unexpected rotations and errors. Despite these challenges, participants demonstrated resilience in navigating difficulties and adapting to the unpredictable nature of the task.
Within the Unpredictable AI condition, participants displayed a diverse range of reactions, spanning from curiosity and enjoyment to stress and frustration. One participant conveyed enthusiasm, expressing a desire to further explore the equipment's potential, saying, "**I want to play with it more. This one was an easy task, I am sure that one can do more or move around more with this equipment.**" Another participant found the task intriguing but noted discomfort, mentioning, "**The task appeared to be simple. I was able to figure out exactly what to do when the task was assigned as the green cube was automatically rotating to indicate the way we have to do the task. But I felt a slight discomfort in vision (eyes) as the images were too close to me.**"
Conversely, a participant described feeling stressed, noting, "**This task was more stressful to complete, as it felt like it was working against me.**" Another participant voiced frustration with the unpredictability of the task, stating, "**In this task, I felt more confident than the one before, and it was kinda easier to know what to do after grabbing the first object. This one was even funnier.**" These quotes highlight the diverse range of emotions and experiences participants encountered when engaging with the Unpredictable AI condition.

5

Discussion

This study aimed to understand how the predictability of AI affects user experience and task performance. The findings are build on the pre-study and pilot study, offering deeper insights into the role of predictability in enhancing or hindering user interactions with AI systems. This section discusses the implications of task completion duration, user satisfaction, control, task load, error rates, and qualitative feedback from participants.

5.1 Task Completion Duration

The significant difference in task completion duration across conditions highlights the impact of AI predictability. Participants completed tasks significantly faster in the predictable condition compared to the manual and unpredictable conditions. This finding aligns with the hypothesis that predictability enhances efficiency in user interactions with AI systems in VR environments. The predictable rotation allowed users to anticipate the system's behavior, reducing cognitive load and enabling quicker task execution. This result underscores the importance of predictability in designing AI systems to streamline user interactions and improve task performance. Predictable systems reduce the time users spend figuring out the system's behavior, leading to quicker task completion and higher productivity. This finding supports existing literature that predictable systems improve task performance (Bubic et al., 2010; Madhavan & Wiegmann, 2007). This is particularly relevant for applications where time efficiency is critical, such as in training simulations, emergency response systems, and productivity tools. Future research could explore how different levels of predictability impact task completion times in more complex and varied tasks.

5.2 User Satisfaction

Despite the significant difference in task completion duration, no significant differences were found in user satisfaction across the different conditions. This suggests that while predictability can enhance efficiency, it does not necessarily translate to higher satisfaction. Participants may value other factors, such as engagement, novelty, or the perceived intelligence of the AI, which were not directly measured in this study. For instance, the novelty of the VR environment or the engagement level of the tasks could play a significant role in overall satisfaction. Additionally, individual differences in user preferences and expectations could influence satisfaction ratings. Further research could explore the relationship between predictability and these additional dimensions of user experience to provide a more comprehensive understanding of how AI system design affects user satisfaction. Daronnat et al. (2021) emphasize that predictability enhances task performance and trust, but this does not directly equate to higher user satisfaction, echoing our observation that predictability improves efficiency but not necessarily satisfaction (Daronnat et al., 2021).

5.3 Control

The sense of control did not significantly differ across the conditions. This indicates that participants felt similarly in control whether they were interacting with the manual, predictable, or unpredictable AI systems. Some participants mitigated perceived unpredictability by manually stopping the cube from rotating freely, which provided a sense of control across all conditions. This behavior might explain the lack of significant differences in perceived control. The ability to intervene allowed participants to manage the cube's rotation, aligning with research suggesting that user-initiated modifications enhance perceived control (Thompson, 2002). This could be due to the VR environment itself, which may inherently provide a strong sense of presence and control. It also suggests that predictability alone does not necessarily enhance the user's perceived control over the interaction. Control is a complex construct that can be influenced by various factors, including the interface design, task complexity, and user familiarity with the system. Future studies could examine how these factors interact with predictability to influence the sense of control. Additionally, exploring control in different VR applications, such as gaming, education, and professional training, could provide more insights into how to design systems that enhance users' perceived control.

5.4 Task Load

The NASA Task Load Index scores showed a marginally significant effect of task condition, indicating that the results approached statistical significance but did not quite reach it. Participants reported lower cognitive load in the predictable condition, which aligns with the faster task completion times observed. This reduced task load in the predictable condition suggests that predictability can make interactions less mentally demanding, thereby making the overall experience more efficient and less tiring for users. A lower task load can lead to better performance, reduced errors, and higher satisfaction in the long term. This finding is supported by Daronnat et al. (2021), who found that agents with more predictable behaviors positively impact cognitive load, reducing the mental effort required to complete tasks. This is crucial for designing AI systems used in high-stakes environments, such as medical procedures, aviation, and complex decision-making scenarios. Future research should explore how different levels of predictability and task complexity interact to influence cognitive load and overall user performance.

5.5 Error Rate

There were no significant differences in error rates across the conditions. This indicates that predictability did not impact the accuracy of task performance. Participants were equally likely to make errors regardless of whether the AI behavior was predictable, unpredictable, or manually controlled. Errors such as grabbing the cube while it was still rotating, knocking the fruits off the table, inserting all the fruits through a single hole, attempting to balance the fruit on the edge instead of fully inserting it, and failing to insert the fruits in the correct order occurred frequently. These errors may have been caused by several factors, including participants being completely new to VR and feeling overwhelmed by the experience, the simulation's real-time effect not feeling entirely realistic, and possibly vague instructions, such as inserting the fruits in a specific order. This result suggests that while predictability can improve efficiency, it does not necessarily affect the accuracy of task execution. Error rates can be influenced by various factors, including task complexity, user experience, and the nature of the errors themselves (e.g., slips versus mistakes). A review in ScienceDirect highlights that while predictability can streamline task performance, it does not necessarily reduce error rates, which supports our findings (Daronnat et al., 2021). Further studies could investigate the types of errors made in different conditions and how predictability might influence specific error patterns.

Understanding these dynamics could help in designing AI systems that not only enhance efficiency but also minimize the likelihood of errors.

5.6 Realistic Expectations for AI Capabilities

The study indicates that while AI predictability can enhance certain aspects of user interaction, such as efficiency and reducing cognitive load, it does not necessarily improve all aspects, such as user satisfaction and error rates. The findings suggest that AI systems, even with high predictability, may not fulfill all the promises of enhancing every dimension of user experience as often advertised. Research shows that predictable AI behaviors can reduce cognitive load and enhance task efficiency by allowing users to anticipate system responses and plan their actions accordingly (Daronnat et al., 2021). However, the impact of AI predictability on user satisfaction and error rates is more nuanced. Factors such as engagement, novelty, and user preferences play significant roles in shaping overall user experience. Studies have shown that while users appreciate the efficiency and reliability of predictable AI, their satisfaction is also influenced by the engagement and novelty the system provides (Bryan-Kinns & Hamilton, 2012; Oh et al., 2016). Additionally, error rates are not necessarily reduced by predictability alone, as other factors like task complexity and user experience also contribute to errors (Daronnat et al., 2021). Engagement and novelty are critical for a positive user experience, and AI systems need to balance predictability with these elements to maintain user interest and satisfaction. Research on recommender systems highlights that diversity and novelty in recommendations can significantly enhance user engagement and satisfaction, beyond mere accuracy (de Gemmis et al., 2015). Furthermore, user preferences vary widely, and what works for one user might not work for another. Studies on adaptive user interfaces emphasize the importance of tailoring AI behaviors to individual user preferences to enhance usability and satisfaction (Halbert & Nathan, 2015). Thus, while AI predictability offers substantial benefits in terms of efficiency and cognitive load reduction, its ability to enhance every aspect of user experience simultaneously is limited. Designers and developers need to consider a holistic approach that includes engagement, novelty, and user preferences to create AI systems that truly enhance the overall user experience.

5.7 Influence of Demographics and the Age Factor

While the age range of our participants (between 18 and 55 years old) was quite broad, it is important to note that we did not include younger participants (under 18) or older participants (over 55). This limitation might affect the generalizability of our findings, as age can influence how users interact with AI-driven systems, particularly in a virtual reality (VR) context. Younger participants, especially those under 18, may exhibit different cognitive and behavioral responses to technology, potentially adapting more quickly to unpredictable AI behaviors due to their increased exposure to digital environments. On the other hand, older adults often experience higher cognitive load and may face more challenges when interacting with unfamiliar or complex technologies like VR. Therefore, while our current sample provides valuable insights, including a more diverse age range in future studies could yield a deeper understanding of how age affects the user experience with AI systems, particularly in terms of predictability, cognitive load, and task performance.

5.8 Other Factors Affecting User Experience

wWhile our study primarily focused on the role of predictability in shaping user experience (UX), it is essential to acknowledge that other factors likely contributed to the overall user interaction with the AI-driven system. Usability, including ease of navigation and clarity of instructions, can greatly affect the user's perception of the system's effectiveness (Nielsen, 1994). Familiarity with technology is another factor to consider; participants with more experience using VR or AI systems may have found the tasks more intuitive, leading to better performance and satisfaction (Venkatesh et al., 2003). Additionally, cognitive load, or the mental effort required to complete a task, plays a critical role in UX, as high cognitive load can reduce user satisfaction and increase errors (Sweller, 1988). Lastly, emotional engagement—how enjoyable or frustrating participants found the tasks—can also affect their overall experience (D. A. Norman, 2004). Addressing these additional UX factors in future research would provide a more comprehensive understanding of user interaction with AI systems, enhancing design and functionality.

5.9 Qualitative Insights

Qualitative feedback provided deeper insights into user experiences. Participants described the predictable condition as "smooth" and "satisfying," whereas the unpredictable condition elicited feelings of frustration and confusion. These qualitative insights highlight the emotional impact of predictability and the importance of designing AI systems that are not only efficient but also enjoyable to use. This emotional response highlights the importance of predictability for overall user experience and should be a key consideration in AI system design (Woźniak et al., 2021). Participants' descriptions of their experiences reveal how predictability can influence their emotional and cognitive responses, which are crucial for long-term acceptance and use of AI systems. For instance, a system that is perceived as predictable and reliable can build trust and confidence, while an unpredictable system can lead to anxiety and reluctance to use the technology. This feedback underscores the value of incorporating qualitative measures in user experience research to capture the nuanced effects of system design on user emotions and perceptions. Future research should include more in-depth qualitative analyses to understand the underlying reasons for user preferences and behaviors.

5.10 Applicability of Findings to Other Settings, Such as Autonomous Driving

The findings of our study, particularly those related to the role of predictability in user experience, have potential applications in other settings, such as autonomous driving. In autonomous vehicles, predictability is crucial for both the driver and passengers, as it ensures that the vehicle behaves in a manner that users can anticipate, thereby fostering trust and comfort (Waytz et al., 2014). Similar to our findings in the virtual reality (VR) environment, where predictable AI behaviors enhanced task efficiency and reduced cognitive load, autonomous driving systems that behave in a predictable and transparent manner are likely to improve user satisfaction and safety (Verberne et al., 2012). For instance, if an autonomous car can clearly signal its next move, such as lane changes or braking, users can better adapt to the system's actions, just as participants in our study adapted to predictable rotations in VR. However, it is important to note that while the parallels between VR-based AI interactions and autonomous driving systems are intriguing, these are still two fundamentally different settings. Autonomous driving involves real-world safety-critical scenarios, whereas VR interactions are controlled experimental environments. Therefore, one must exercise caution when drawing direct conclusions from our findings to such high-stakes contexts. More research would be required to fully understand how predictability impacts user experience in autonomous driving, considering the different variables and risks involved.

5.11 VR Experience

Working with VR as our project environment was a completely new experience, offering significant learning opportunities, from building the prototype in Unity to controlling the environment and seeing participants enjoy the study. However, the experience wasn't entirely positive. At times, we felt frustrated by the complexity of the setup, such as connecting the VR headset to the PC, whether using a cable or wirelessly. Even connecting the headset to the university Wi-Fi took time, and we had limited resources for troubleshooting any issues that arose. Additionally, working continuously with the VR headset sometimes caused dizziness, adding to the challenges we faced.

5.12 Ethical Considerations

The study adhered to ethical principles outlined by relevant institutional review boards and guidelines for research involving human participants. Informed consent was obtained from all participants, ensuring voluntary participation and the right to withdraw at any time without repercussion. Anonymity and confidentiality were maintained, with data stored securely and accessible only to authorized personnel. Measures were taken to minimize potential risks or discomfort to participants during data collection and experimental procedures. All research activities were conducted in accordance with applicable laws and regulations governing research ethics. Any potential conflicts of interest were disclosed and managed appropriately.

5.13 Limitations

One notable limitation is the potential influence of participants' excitement about using VR for the first time, which may have positively biased their overall experience. Additionally, the study focused on a specific task within a VR environment, which may not fully capture the complexity of real-world AI interactions. Extending the research to various tasks and settings could provide a more comprehensive understanding of predictability in AI systems."

5.14 Future Research

Future research should explore several avenues to deepen our understanding of the role of predictability in AI systems and its impact on user experience:

1. Diverse AI Applications: Investigate predictability in various AI applications beyond virtual reality, such as autonomous vehicles, personal assistants, and healthcare systems, to understand broader implications in different contexts.
2. Longitudinal Studies: Conduct longitudinal studies to assess how user satisfaction and performance evolve over time with continuous interaction with increased complexity of predictable and unpredictable AI systems. These studies could provide valuable information on the long-term effects of AI predictability on user experience since they would be using the system for a certain time so both positive and negative experience might occur.
3. Diverse Participant Sample: Include a more diverse participant sample in terms of age, background, and technical expertise to ensure the findings are generalizable to a broader population. Understanding diverse needs and preferences is crucial for developing inclusive and effective AI systems.
4. Advanced Predictability Models: Develop and test sophisticated models of predictability that can adapt to individual user preferences and contexts, enhancing personalization in AI interactions. Tailored experiences could lead to higher levels of user satisfaction and engagement.

6

Conclusion

This research aimed to explore the influence of predictability in AI-driven virtual reality (VR) environments on user experience and task performance. Our findings underscore the pivotal role of predictability in enhancing task efficiency and user interactions within VR settings. Participants performed tasks significantly faster in the predictable condition compared to both manual and unpredictable conditions, highlighting the benefits of predictable AI behavior for streamlining interactions and improving performance.

Despite these improvements in task completion times, our quantitative measures did not reveal significant differences in user satisfaction, perceived control, NASA Task Load Index (NASA-TLX) scores, or error rates across the different conditions. This suggests that while predictability enhances efficiency, it may not substantially affect other aspects of user experience. However, qualitative feedback provided deeper insights, revealing that participants found the predictable condition to be "smooth" and "satisfying," whereas the unpredictable condition elicited frustration and confusion. These emotional responses indicate that predictability contributes to a more positive user experience and should be a key consideration in AI system design.

The study acknowledges the potential influence of the novelty effect of VR, which might have positively biased participants' overall experience. Furthermore, the focus on a specific task within a controlled VR environment may limit the generalizability of our findings to real-world AI applications. Future research should aim to investigate predictability in diverse AI applications and conduct longitudinal studies to assess how user satisfaction and performance evolve over time with continuous interaction with AI systems.

Including a more diverse participant sample and developing sophisticated models of predictability tailored to individual user preferences and contexts could offer deeper insights into the nuances of user experiences and enhance personalization in AI interactions. Such efforts will contribute to a more thorough understanding of how predictability influences user experience and help design AI systems that are more intuitive, personalized, and effective across various contexts.

In conclusion, this research contributes to the evolving understanding of predictability in AI systems and underscores the importance of considering user experiences from both quantitative and qualitative perspectives. As AI continues to permeate various aspects of daily life, ongoing research in this area is essential for fostering positive interactions between humans and intelligent systems.

Bibliography

Bergström, J., Dalsgaard, T.-S., Alexander, J., & Hornbæk, K. (2021). How to Evaluate Object Selection and Manipulation in VR? Guidelines from 20 Years of Studies. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–20. https://doi.org/10.1145/3411764.3445193

Boffetta, G., Cencini, M., Falcioni, M., & Vulpiani, A. (2002). Predictability: A way to characterize complexity. *Physics Reports*, *356*(6), 367–474. https://doi.org/10.1016/S0370-1573(01)00025-4

Bryan-Kinns, N., & Hamilton, F. (2012). Identifying mutual engagement [Publisher: Taylor & Francis]. *Behaviour & Information Technology*, *31*(2), 101–125. https://doi.org/10.1080/01449290903377103

Bubic, A., von Cramon, D. Y., & Schubotz, R. I. (2010). Prediction, Cognition and the Brain. *Frontiers in Human Neuroscience*, *4*, 25. https://doi.org/10.3389/fnhum.2010.00025

Chougale, P., Yadav, V., Gaikwad, A., Student, B., & Vidyapeeth. (2022). FIREBASE -OVERVIEW AND USAGE. *Journal of Engineering and Technology Management*, 2582–5208.

Clark, I. A., & Maguire, E. A. (2020). Do questionnaires reflect their purported cognitive functions? *Cognition*, *195*, 104114. https://doi.org/10.1016/j.cognition.2019.104114

Clodic, A., Pacherie, E., Alami, R., & Chatila, R. (2017). Key Elements for Human-Robot Joint Action. In R. Hakli & J. Seibt (Eds.), *Sociality and Normativity for Robots: Philosophical Inquiries into Human-Robot Interactions* (pp. 159–177). Springer International Publishing. https://doi.org/10.1007/978-3-319-53133-5_8

Daronnat, S., Azzopardi, L., Halvey, M., & Dubiel, M. (2021). Inferring Trust From Users' Behaviours; Agents' Predictability Positively Affects Trust, Task Performance and Cognitive Load in Human-Agent Real-Time Collaboration [Publisher: Frontiers Media S.A.]. *Frontiers in Robotics and AI*, *8*. https://doi.org/10.3389/frobt.2021.642201 de Gemmis, M., Lops, P., Semeraro, G., & Musto, C. (2015). An investigation on the serendipity problem in recommender systems. *Information Processing & Management*, *51*(5), 695–717. https://doi.org/10.1016/j.ipm.2015.06.008

de Vreede, G.-J., Fruhling, A., & Chakrapani, A. (2005). A Repeatable Collaboration Process for Usability Testing [ISSN: 1530-1605]. *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*, 46–46. https://doi.org/10.1109/HICSS.2005.46

Fernando, T., & Hodrien, A. (2021, May). A Review of Post-Study and Post-Task Subjective Questionnaires to Guide Assessment of System Usability - JUX. Retrieved June 19, 2024, from https://uxpajournal.org/review-post-study-task-subjective-questionnaires-usability/

Groover, M. P. (2020, July). *Fundamentals of Modern Manufacturing: Materials, Processes, and Systems* [Google-Books-ID: mB7zDwAAQBAJ]. John Wiley & Sons.

Halbert, H., & Nathan, L. P. (2015). Designing for Discomfort: Supporting Critical Reflection through Interactive Tools. *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 349–360. https://doi.org/10.1145/2675133.2675162

Hoff, K. A., & Bashir, M. (2015). Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust [Publisher: SAGE Publications Inc]. *Human Factors*, *57*(3), 407–434. https://doi.org/10.1177/0018720814547570

Johnson, J. (2020, June). 4 Types of Artificial Intelligence. Retrieved February 20, 2024, from https://www.bmc.com/blogs/artificial-intelligence-types/

Johnson, R., & Onwuegbuzie, A. (2004). Mixed Methods Research: A Research Paradigm Whose Time Has Come. *Educational researcher*, *33*, 14. https://doi.org/10.3102/0013189X033007014

Kim, T., & Hinds, P. (2006). Who Should I Blame? Effects of Autonomy and Transparency on Attributions in Human-Robot Interaction [ISSN: 1944-9437]. *RO-MAN 2006 - The 15th IEEE International Symposium on Robot and Human Interactive Communication*, 80–85. https://doi.org/10.1109/ROMAN.2006.314398

Kuang, E. (2023). Crafting Human-AI Collaborative Analysis for User Experience Evaluation. *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–6. https://doi.org/10.1145/3544549.3577042

Lee, J. D., & See, K. A. (2004). Trust in Automation: Designing for Appropriate Reliance [Publisher: SAGE Publications Inc]. *Human Factors*, *46*(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392

Li, T., Vorvoreanu, M., Debellis, D., & Amershi, S. (2023). Assessing Human-AI Interaction Early through Factorial Surveys: A Study on the Guidelines for Human-AI Interaction. *ACM Transactions on Computer-Human Interaction*, *30*(5), 69:1–69:45. https://doi.org/10.1145/3511605

Madhavan, P., & Wiegmann, D. A. (2007). Similarities and differences between human–human and human–automation trust: An integrative review [Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/14639220500337708]. *Theoretical Issues in Ergonomics Science*, *8*(4), 277–301. https://doi.org/10.1080/14639220500337708

Newell, A., & Simon, H. (1956). The logic theory machine–A complex information processing system [Conference Name: IRE Transactions on Information Theory]. *IRE Transactions on Information Theory*, *2*(3), 61–79. https://doi.org/10.1109/TIT.1956.1056797

Nielsen, J. (1994, September). *Usability Engineering* [Google-Books-ID: 95As2OF67f0C]. Morgan Kaufmann.

Norman, D. A. (2004). Emotional Design: Why We Love (or Hate) Everyday Things - ProQuest. Retrieved October 1, 2024, from https://www.proquest.com/openview/c58ae698d218e5b0fe68cabd0aa3b637/1?pq-origsite=gscholar&cbl=29587

Norman, D. (2014, January). The Design of Everyday Things. Retrieved June 12, 2024, from https://mitpress.mit.edu/9780262525671/the-design-of-everyday-things/

Obhi, S. S., & Hall, P. (2011). Sense of agency and intentional binding in joint action. *Experimental Brain Research*, *211*(3), 655–662. https://doi.org/10.1007/s00221-011-2675-2

Oh, S. Y., Bailenson, J., Krämer, N., & Li, B. (2016). Let the Avatar Brighten Your Smile: Effects of Enhancing Facial Expressions in Virtual Environments [Publisher: Public Library of Science]. *PLOS ONE*, *11*(9), e0161794. https://doi.org/10.1371/journal.pone.0161794

Onwuegbuzie, A. J., Dickinson, W. B., Leech, N. L., & Zoran, A. G. (2009). A Qualitative Framework for Collecting and Analyzing Data in Focus Group Research [Publisher: SAGE Publications Inc]. *International Journal of Qualitative Methods*, *8*(3), 1–21. https://doi.org/10.1177/160940690900800301

Protopsaltis, A., & Papagiannakis, G. (2020). Virtual Reality Systems, Tools, and Frameworks. In N. Lee (Ed.), *Encyclopedia of Computer Graphics and Games* (pp. 1–6). Springer International Publishing. https://doi.org/10.1007/978-3-319-08234-9_102-1

Russell, S., & Norvig, P. (2020). *Artificial Intelligence: A Modern Approach* (4th Edition). Pearson. https://www.pearson.com/en-us/subject-catalog/p/artificial-intelligence-a-modern-approach/P200000003500/9780134610993?tab=title-overview

Rzepka, C., & Berger, B. (2018, December). *User Interaction with AI-enabled Systems: A Systematic Review of IS Research.*

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences*, *10*(2), 70–76. https://doi.org/10.1016/j.tics.2005.12.009

Strother, L., House, K. A., & Obhi, S. S. (2010). Subjective agency and awareness of shared actions. *Consciousness and Cognition*, *19*(1), 12–20. https://doi.org/10.1016/j.concog.2009.12.007

Swan, M. B., & Notess, M. (2003). Predicting user satisfaction from subject satisfaction. *CHI '03 Extended Abstracts on Human Factors in Computing Systems*, 738–739. https://doi.org/10.1145/765891.765960

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–285. https://doi.org/10.1016/0364-0213(88)90023-7

Thompson, S. C. (2002). The role of personal control in adaptive functioning. In *Handbook of positive psychology* (pp. 202–213). Oxford University Press.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User Acceptance of Information Technology: Toward a Unified View [Publisher: Management Information Systems Research Center, University of Minnesota]. *MIS Quarterly*, *27*(3), 425–478. https://doi.org/10.2307/30036540

Verberne, F. M. F., Ham, J., & Midden, C. J. H. (2012). Trust in Smart Systems: Sharing Driving Goals and Giving Information to Increase Trustworthiness and Acceptability of Smart Systems in Cars [Publisher: SAGE Publications Inc]. *Human Factors*, *54*(5), 799–810. https://doi.org/10.1177/0018720812443825

Waytz, A., Heafner, J., & Epley, N. (2014). The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology*, *52*, 113–117. https://doi.org/10.1016/j.jesp.2014.01.005

Weistroffer, V., Paljic, A., Callebert, L., & Fuchs, P. (2013). A methodology to assess the acceptability of human-robot collaboration using virtual reality. *Proceedings of the 19th ACM Symposium on Virtual Reality Software and Technology*, 39–48. https://doi.org/10.1145/2503713.2503726

Wen, W., & Haggard, P. (2018). Control Changes the Way We Look at the World. *Journal of Cognitive Neuroscience*, *30*(4), 603–619. https://doi.org/10.1162/jocn_a_01226

Wen, W., & Imamizu, H. (2022). The sense of agency in perception, behaviour and human–machine interactions [Publisher: Nature Publishing Group]. *Nature Reviews Psychology*, *1*(4), 211–222. https://doi.org/10.1038/s44159-022-00030-6

Woźniak, P. W., Karolus, J., Lang, F., Eckerth, C., Schöning, J., Rogers, Y., & Niess, J. (2021). Creepy Technology:What Is It and How Do You Measure It? *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–13. https://doi.org/10.1145/3411764.3445299

Yang, Q., Steinfeld, A., Rosé, C., & Zimmerman, J. (2020). Re-examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–13. https://doi.org/10.1145/3313831.3376301

A

Study Design Document

A.1 Purpose of study (research hypothesis/questions)

The purpose of this study is to answer the following research question:

> *Does User predictability affect user satisfaction in Human-AI interaction within Virtual Reality environments?*

Hypotheses:

- **H0.1**: Predictability does not significantly affect user satisfaction and engagement in Human-AI interactions within Virtual Reality environments.
- **HA.1**: Predictability significantly affects user satisfaction and engagement in Human-AI interactions within Virtual Reality environments.

A.2 Participant profile

Participants should (not) have the following criteria:

- Should not be color blind.
- Should have normal or corrected-to-normal vision.
- Should not have impaired hand function.
- Should be an adult (18 years or older).

A.3 Compensation plan

There will not be a compensation for this study.

A.4 Methodology

A.4.1 Study Design

The study will follow a mixed design:

- **Quantitative data** collected:
  - Time to complete each of the conditions in the task.
  - Error rate.
  - Eye tracking.
- **Qualitative data** collected via open text fields in questionnaires:
  - Participants will describe how the interaction with the different conditions made them feel (their satisfaction).

Summative: There will not be any follow-up after conducting the study.

A.4.2 Experiments

- **Design**: Within-subject/repeated measures design; participants will participate and evaluate all conditions.
- **Estimated number of participants**: 20.
- **Conditions**:
  1. **Active Condition**: No help or assistance from the system.
  2. **AI Predicted Assistant**: Participants receive help/assistance from the system, which behaves in a predictable way.
  3. **AI Un-Predicted Assistant**: Participants receive help/assistance from the system, which behaves in an unpredictable way.

A.4.3 Variables

- **Independent variables**: Predictability.
- **Dependent variables**: Satisfaction (or SUS), Time.
- **Confounding variables/Covariates**: Being-in-control, Agency.

A.5 Material & Apparatus

- **Research artifacts**: Meta Quest Pro Premium mixed reality, 3D-Unity based environment, one computer, and two monitors.
- **3D-Unity based environment**:
  - One cube with 4 different engraved faces.
  - 4 different fruits, each fitting into its corresponding engraved face.
- **Questionnaire**:
  - General demographic questions (such as gender and age).
  - Questions about user satisfaction and interaction.
- **Informed consent form**, including study instructions.

- **Tasks**:
  - Reading the consent form before starting the task.
  - Filling out general information about themselves.
  - Signing the consent form.
  - Completing the pre-study phase to familiarize themselves with the VR environment.
  - After each condition, participants will fill out a questionnaire evaluating their satisfaction and being-in-control.

A.6 Structure & Timeline

A.6.1 Preparation required before the study

- Consent form.
- Demographic data (filled in consent form).
- Assessment of influential factors (study questionnaire), including:
  - Experience with assistive technology and VR.
  - Age.
  - Color blindness.
  - Highest education level.
  - Gender.

A.6.2 Timeline of study phase

- **Subject**: Time allocation for each task.

Task	**Time**
Participant reads and signs the consent form.	1-2 minutes
Participant fills in demographic information.	2nd to 3rd minute
Participant puts on the VR set.	4th to 6th minute
Participant engages with condition 1 and answers the questionnaire.	7th to 14th minute
Short break (optional).	14th to 15th minute
Participant engages with condition 2 and answers the questionnaire.	15th to 18th minute
Short break (optional).	18th to 19th minute
Participant engages with condition 3 and answers the questionnaire.	19th to 22nd minute

A.7 Analysis

A.7.1 Quantitative Analysis

- **RStudio**: For data import, access, transformation, exploration, plotting, and machine learning.
- **Repeated Measures ANOVA**: To analyze differences in task completion time, satisfaction, and being-in-control across the conditions.
- **Tukey**: For post-hoc analysis if significant effects are found in the ANOVA.
- **Correlation Analysis**: To explore relationships between task completion time, satisfaction, and being-in-control.

A.7.2 Qualitative Analysis

- **Atlas.ti**: For analysis of qualitative data.
- **Thematic Analysis**: To identify and report themes related to user experiences and satisfaction.
- **Content Analysis**: To quantify the presence and meaning of words and concepts in qualitative responses.
- **Comparative Analysis**: To compare feedback across conditions regarding the predictability of AI assistance.

A.7.3 Data Forms/Logs

- Timer.
- Recorded eye-tracking data.
- Raw data (CSV) from the questionnaire.

A.8 Personnel Involved & Responsibilities

- **Ismael & Ishwor**: Preparation of documents (consent form, questionnaire), design of VR environment.
- **Ishwor**: Prototype finalization.
- **Ismael**: Setup study design/methodology.
- **Ismael & Ishwor (and possibly Paweł)**: Recruitment of participants.
- **Paweł**: Feedback and research approval.

B

Consent Form

Informed Consent Form

Title of Study: Assessing the impact of predictability on Interaction with AI in VR.

Researchers: Ishwor & Ismael

Purpose of the Study

The purpose of this study is to investigate the impact of different rotational conditions on object manipulation within a virtual reality (VR) environment. Specifically, we aim to examine the predictability of object manipulation when the cube rotates along specific axes compared to random rotations.

Procedure

If you agree to participate, you agree to:

- **Wear a VR headset and interact with the prototype VR system.**
- **Complete a series of tasks within the VR environment.**
- **Answer a questionnaire about your experience using the VR system.**
- **Optionally, participate in a follow-up interview to provide further feedback.**
- **Must have normal or corrected to normal vision.**
- **Must not have the impaired hand function.**

Risks and Benefits

There are no known physical risks associated with participation in this study. However, some participants may experience mild discomfort or motion sickness due to the VR environment. Benefits include contributing to research aimed at improving understanding of VR interaction and potentially enhancing VR user experiences in the future.

Confidentiality

All data collected will be kept confidential and will only be accessible to the research team. Participant identities will remain anonymous in any publications or reports resulting from this study.

Voluntary Participation

Participation in this study is entirely voluntary. Participants are free to withdraw from the study at any time without penalty or consequence.

Statement of Consent

I have read and understand the information provided above. I voluntarily agree to participate in this study and consent to the use of my data for research purposes.
Participant's Signature:

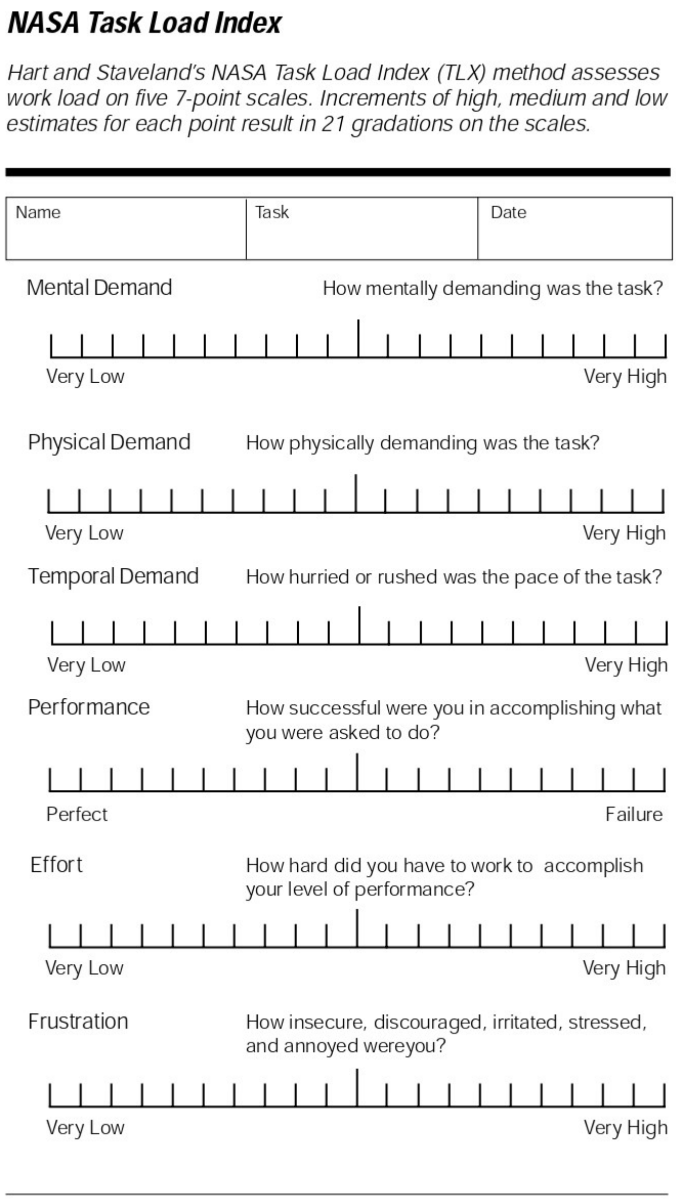
Date:

By signing this form, you indicate your understanding of the study procedures and your voluntary agreement to participate.

C

Nasa-TLX Form

Figure C.1: NASA-TLX



NASA Task Load Index

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

Name	Task	Date

Mental Demand — How mentally demanding was the task?

Very Low — Very High

Physical Demand — How physically demanding was the task?

Very Low — Very High

Temporal Demand — How hurried or rushed was the pace of the task?

Very Low — Very High

Performance — How successful were you in accomplishing what you were asked to do?

Perfect — Failure

Effort — How hard did you have to work to accomplish your level of performance?

Very Low — Very High

Frustration — How insecure, discouraged, irritated, stressed, and annoyed wereyou?

Very Low — Very High