

Chair for Computer Science 10 (Media Computing and Human-Computer Interaction)

RWTH AACHEN UNIVERSITY

Viewport and Window Management in Augmented Reality

Master's Thesis at the
Media Computing Group
Prof. Dr. Jan Borchers
Computer Science Department
RWTH Aachen University

by
Martin Zakrzewski

Thesis advisor:
Prof. Dr. Jan Borchers

Second examiner:
Prof. Dr. Ulrik Schroeder

Registration date: 12.01.2026
Submission date: 12.08.2026

Zentrales Prüfungsamt/Central Examination Office

Eidesstattliche Versicherung
Declaration of Academic Integrity

Zakrzewski, Martin
Name, Vorname/Last Name, First Name

440557
Matrikelnummer (freiwillige Angabe)
Student ID Number (optional)

Ich versichere hiermit an Eides Statt, dass ich die vorliegende ~~Arbeit/Bachelorarbeit/~~ Masterarbeit* mit dem Titel
I hereby declare under penalty of perjury that I have completed the present paper/bachelor's thesis/master's thesis* entitled

Viewport and Window Management in Augmented Reality

selbstständig und ohne unzulässige fremde Hilfe (insbes. akademisches Ghostwriting) erbracht habe. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt; dies umfasst insbesondere auch Software und Dienste zur Sprach-, Text- und Medienproduktion. Ich erkläre, dass für den Fall, dass die Arbeit in unterschiedlichen Formen eingereicht wird (z.B. elektronisch, gedruckt, geplottet, auf einem Datenträger) alle eingereichten Versionen vollständig übereinstimmen. Die Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.
independently and without unauthorized assistance from third parties (in particular academic ghostwriting**.** I have not used any other sources or aids than those indicated; this includes in particular software and services for language, text, and media production. In the event that the work is submitted in different formats (e.g. electronically, printed, plotted, on a data carrier), I declare that all the submitted versions are fully identical. I have not previously submitted this work, either in the same or a similar form to an examination body.

Aachen, 12.08.2026
Ort, Datum/City, Date


Unterschrift/Signature

*Nichtzutreffendes bitte streichen/Please delete as appropriate

Belehrung:
Official Notification:

§ 156 StGB: Falsche Versicherung an Eides Statt
Wer vor einer zur Abnahme einer Versicherung an Eides Statt zuständigen Behörde eine solche Versicherung falsch abgibt oder unter Berufung auf eine solche Versicherung falsch aussagt, wird mit Freiheitsstrafe bis zu drei Jahren oder mit Geldstrafe bestraft.
§ 156 StGB (German Criminal Code): False Unsworn Declarations
Whosoever before a public authority competent to administer unsworn declarations (including Declarations of Academic Integrity) falsely submits such a declaration or falsely testifies while referring to such a declaration shall be liable to imprisonment for a term not exceeding three years or to a fine.
§ 161 StGB: Fahrlässiger Falscheid; fahrlässige falsche Versicherung an Eides Statt
(1) Wenn eine der in den §§ 154 bis 156 bezeichneten Handlungen aus Fahrlässigkeit begangen worden ist, so tritt Freiheitsstrafe bis zu einem Jahr oder Geldstrafe ein.
(2) Straflosigkeit tritt ein, wenn der Täter die falsche Angabe rechtzeitig berichtigt. Die Vorschriften des § 158 Abs. 2 und 3 gelten entsprechend.
§ 161 StGB (German Criminal Code): False Unsworn Declarations Due to Negligence
(1) If an individual commits one of the offenses listed in §§ 154 to 156 due to negligence, they are liable to imprisonment for a term not exceeding one year or to a fine.
(2) The offender shall be exempt from liability if they correct their false testimony in time. The provisions of § 158 (2) and (3) shall apply accordingly.

Die vorstehende Belehrung habe ich zur Kenntnis genommen:
I have read and understood the above official notification:

Aachen, 12.08.2026
Ort, Datum/City, Date


Unterschrift/Signature

Contents

Abstract xiii

Überblick xv

Acknowledgments xvii

Conventions xix

1 Introduction 1

1.1 Window Management and Constraints 3

1.2 Aim and Research Questions 5

1.3 Outline 7

2 Related Work 9

2.1 Window Management on Desktop Displays 9

2.2 Beyond the Single Screen 11

2.3 Window and Content Management in AR 13

2.4 Mental Models 15

2.5 Low-Fidelity Prototyping 16

2.6 Summary 17

3 Eliciting Mental Models of AR Window Management 19

3.1 Methodology 20

3.1.1 Study Goals and Approach 20

3.1.2 Study Design and Conditions 21

3.1.3 Apparatus: Designing the Low Fidelity Prototype 23

3.1.4 Task and Scenario 26

3.1.5 Participants and Procedure 28

3.1.6 Measures and Analysis 29

3.2 Results and Evaluation 32

3.2.1 Demographics 32

3.2.2 Quantitative Results 33

3.2.3 Qualitative Results 34

T1: Importance Influences Size, Place, and Visibility 35

T2: Window Organization 38

T3: Bringing Windows into View 41

T4: Using Desktop Paradigms 45

T5: Freedom vs Structure 46

3.2.4 Prototype Limitations 48

3.3 Summary of Findings 50

3.4 Limitations 52

4 The High-Fidelity Prototype **55**

4.1 From Findings to Design 55

4.1.1 From Horizontal Constraint to Free Space 56

4.1.2 Conditions as a Response to Freedom and Structure (T5) . . . 56

4.1.3 Window Behavior (T1) 58

4.1.4 Grouping and Overlap (T2) 59

4.1.5 Bringing Windows into View (T3) 60

4.1.6 Desktop Familiarity (T4) 62

4.2 Implementation 62

5 Prototype Evaluation **67**

5.1 Methodology 67

5.1.1 Study Design and Conditions 68

5.1.2 Apparatus and Setting 69

5.1.3 Task and Scenario 69

5.1.4 Participants and Procedure 71

5.1.5 Measures and Analysis 72

5.2 Results 74

5.2.1 Manipulation Check 74

5.2.2 Selection Accuracy (H1) 75

5.2.3 Search Time (H2a, H2b) 76

5.2.4 Perceived Workload (H3) 77
5.2.5 Usability (H4) 77
5.2.6 Layout Ratings 78
5.2.7 Rankings and Order Effects 79
5.3 Qualitative Results 79
5.3.1 Grouping as the preferred structure 80
5.3.2 Freedom as the Baseline Expectation 82
5.3.3 The rejection of the scrolling mechanism 84
5.3.4 The demand for flexibility inside structure 85
5.3.5 The persistence of desktop concepts 86
5.3.6 Implementation Feedback 87
5.4 Discussion 88

6 Summary and Future Work **95**
6.1 Summary and Contributions 95
6.2 Limitations and Future Work 97

A Questionnaires Study 1 **101**

B Codebook Study 1 **111**

C Drawings Study 1 **115**

D Questionnaires Study 2 **121**

E Codebook Study 2 **131**

Bibliography **135**

Index **141**

List of Figures and Tables

3.1 *Low Fidelity Prototype:* Prototype positioned on the table 24

3.2 *Window Representations:* From macOS window to prototype card . . . 26

3.3 *Demographics:* Participants of the Formative Study 32

3.4 *Questionnaire Ratings:* Condition ratings from the prototype study . . 34

3.5 *Likert Distributions:* Response distributions per item and condition . . 35

3.6 *Rankings:* Condition rankings per task category 36

3.7 *Themes:* The five themes from the reflexive thematic analysis 37

3.8 *Elastic Horizontal Layout:* Windows in three sizes on the strip 37

3.9 *Free Placement Layout:* Thematic clusters separated by gaps 40

3.10 *Fixed Horizontal Layout:* The constrained strip with unused space . . . 48

3.11 *Co-Design Drawings:* Visual patterns matching the elicited mental models 49

4.1 *Free Placement:* Unbounded arrangement with three window sizes . . 57

4.2 *Grouped Placement:* User-created groups with bounding boxes 58

4.3 *Scrollbar:* Groups extended with a scrolling mechanism 59

4.4 *Snap Preview:* Red insertion indicator and translucent dragging . . . 60

4.5 *Taskbar:* Window access and study controls 63

4.6 *Study Layout:* A full 20-window workspace during the search task . . 66

5.1 *Demographics:* Participants of the Evaluation Study 72

5.2 *Scroll Usage:* Scrollbar-mechanic usage per participant 75

5.3 *Search Times:* Per-participant median search times per condition . . . 76

5.4 *SUS Scores:* Usability ratings per condition 78

A.1 *Questionnaire Study 1:* Demographics 102

A.2 *Questionnaire Study 1:* Stop warning between parts 103

A.3 *Questionnaire Study 1:* Ratings, Free Placement 104

A.4 *Questionnaire Study 1:* Ratings, Fixed Horizontal 105

A.5 *Questionnaire Study 1:* Ratings, Elastic Horizontal 106

A.6 *Questionnaire Study 1:* Overall ranking 107

A.7 *Questionnaire Study 1:* Ranking by task type 108

A.8 *Questionnaire Study 1:* Ranking, communication and email 109

D.1 *Questionnaire Study 2:* Demographics 122

D.2 *Questionnaire Study 2:* Layout ratings, conditions and first items . . . 123

D.3 *Questionnaire Study 2:* Layout ratings, remaining items 124

D.4 *Questionnaire Study 2:* NASA RawTLX, first dimensions 125

D.5 *Questionnaire Study 2:* NASA RawTLX, remaining dimensions 126

D.6 *Questionnaire Study 2:* System Usability Scale 127

D.7 *Questionnaire Study 2:* Usability items and open feedback 128

D.8 *Questionnaire Study 2:* Final ranking 129

D.9 *Questionnaire Study 2:* Final reflections 130

Abstract

Head-worn displays such as *Apple Vision Pro* remove the physical screen that has structured window management for decades. Content can be placed anywhere in a 360° space, free of edges, size limits, and monitor counts. However, most Augmented Reality window and viewport management systems reintroduce a bounded, monitor-like region, keeping the constraints of the desktop without asking whether users want them. While there has been extensive research on interaction techniques for arranging content in AR, evidence of what users expect such a system to do in the first place remains sparse. To ground the design of AR window management in what users actually want, we needed to gain insight into their mental models of the space. In our work, we investigated the following research questions: 1. What mental model do users have of window and viewport management in AR? 2. When arranging windows, how much and what kind of constraint do users want? 3. Do the constrained conditions improve the usability of window management compared to the free placement baseline? 4. Which window management approach best matches users' mental models and preferences? We addressed these questions in two studies. In a formative study, 24 participants arranged window cards on a low-fidelity prototype across three conditions with different constraints. Interviews and a co-design drawing task followed this activity. Our findings suggest that users organize the space by importance and meaning, retrieve windows by remembering where they placed them, and carry desktop conventions into AR. Participants rejected constraints imposed on them but welcomed structure they chose themselves. We translated these findings into SpatialWindowPlacer, a prototype for *Apple Vision Pro*. For this prototype, we created three designs and evaluated them across 12 participants. Search time, selection accuracy, and workload did not differ between the designs. User-instantiated grouping carried no penalty to perceived control and was preferred by a majority, while a scrolling mechanism derived from participants' own descriptions was rejected in use. Our work suggests that constraints inherited from the physical screen should be offered to users rather than imposed on them.

Überblick

Kopfgetragene Displays wie das *Apple Vision Pro* ersetzen den physischen Bildschirm, der jahrzehntelang die Fensterverwaltung geprägt hat. Inhalte können überall in einem 360°-Raum platziert werden, frei von Rahmen, Größenbeschränkungen und Monitoranzahl. Jedoch führen die meisten AR-Fenster- und Viewport-Management-Systeme wieder eine begrenzte, monitorähnliche Region ein, die die Einschränkungen des Desktops beibehält, ohne zu fragen, ob die Benutzer diese Einschränkungen wollen. Obwohl es umfangreiche Forschung zu Interaktionstechniken für die Anordnung von Inhalten in AR gibt, sind Untersuchungen darüber, was Benutzer von einem solchen System überhaupt erwarten, immer noch spärlich. Um das Design des AR-Fenstermanagements auf das zu stützen, was die Nutzer tatsächlich wollen, mussten wir Einblicke in ihre mentalen Modelle zur AR-Fenstergestaltung gewinnen. In unserer Arbeit haben wir die folgenden Forschungsfragen untersucht: 1. Welches mentale Modell haben Benutzer von Fenster- und Viewport-Management in AR? 2. Wie viel und welche Art von Einschränkung wünschen sich die Benutzer bei der Anordnung von Fenstern? 3. Verbessern eingeschränkte Bedingungen die Benutzerfreundlichkeit des Fenstermanagements im Vergleich zur freien Platzierung? 4. Welcher Ansatz des Fenstermanagements entspricht den mentalen Modellen und Vorlieben der Benutzer am besten? Wir haben dies in zwei Studien behandelt. In einer formativen Studie ordneten 24 Teilnehmer Fensterkarten auf einem Low-Fidelity-Prototyp in drei Bedingungen mit unterschiedlichen Einschränkungen. Danach folgten Interviews und eine Co-Design-Zeichnungsaufgabe. Unsere Ergebnisse deuten darauf hin, dass Benutzer den Raum nach Wichtigkeit und Bedeutung organisieren, Fenster abrufen, indem sie sich daran erinnern, wo sie diese platziert haben, und Desktop-Konventionen in AR übertragen. Die Teilnehmer lehnten auferlegte Einschränkungen ab, begrüßten jedoch die von ihnen selbst gewählte Struktur. Wir haben diese Erkenntnisse in SpatialWindowPlacer, einem Prototyp für die *Apple Vision Pro*, übertragen. Für diesen Prototypen haben wir drei Designs erstellt und diese an 12 Teilnehmern evaluiert. Suchzeit, Auswahlgenauigkeit und Arbeitsbelastung unterschieden sich unter den Designs nicht. Die vom Benutzer initiierte Gruppierung hatte keine negativen Auswirkungen auf die wahrgenommene

Kontrolle und wurde von der Mehrheit bevorzugt, während ein aus den eigenen Beschreibungen der Teilnehmer abgeleiteter Scrollmechanismus in der Anwendung abgelehnt wurde. Unsere Arbeit legt nahe, dass Einschränkungen, die vom physischen Bildschirm stammen, den Nutzern angeboten und nicht aufgezwungen werden sollten.

Acknowledgments

First of all, I would like to thank Prof. Dr. Jan Borchers and Prof. Dr. Ulrik Schroeder for examining my thesis.

Many thanks to Kevin Fiedler and René Schäfer for supervising my work! I have learned a lot thanks to both of you and your valuable feedback.

I also want to thank my family, my girlfriend, and my friends for being such amazing support. I will not name anyone individually, in case I forget someone. Sorry!

Moreover, another huge thank you to the people at i10. Without your feedback and help, this thesis would not have been possible.

Finally, thank you to everyone who took the time to participate in my studies, and to those who came back for the second one. Your patience with the laminated cards and the *Apple Vision Pro* prototype made this work possible.

Again, thank you to all of you!

P.S. Thank you, Yangmal, for all the emotional support!



This is Yangmal!

Conventions

Throughout this thesis we use the following conventions:

- The thesis is written in American English.
- The first person is written in plural form.
- Unidentified third persons are referred neutrally or in the plural form.

Definitions of technical terms are set off in coloured boxes.

DEFINITION:
Definitions are set off in orange boxes.

Where appropriate, paragraphs are summarized by one or two sentences that are positioned at the margin of the page.

This is a summary of a paragraph.

Use of AI Tools: In this thesis AI tools where used for writing/coding support, translation and proof reading.

Chapter 1

Introduction

HWDs can give a 360° space for arranging information.

Head-worn displays (HWDs) such as the *Apple Vision Pro*, *Meta Quest*, and *Valve Index* headsets no longer tie windows to a physical screen: Content can be placed anywhere in a 360° environment. This functionality removes the edges, size limits, and number of available monitors that have traditionally structured window management. Using the full 360° range of HWDs does not come without cost. Previous work has shown that arranging and viewing content in such a wide field requires large, repeated head movements, which is linked to neck and back strain [Grubert et al., 2018]. In practice, many Augmented Reality (AR) window management systems do not use the full surrounding space. They anchor content to a bounded, monitor-like region at a fixed position [Pavanatto et al., 2025a,b]. These systems reintroduce constraints by design, primarily to keep content within comfortable reach. However, this design decision is made for users, not by them. Once the physical screen is gone and the entire space becomes available, it remains open how much constraint, and what kind, users themselves actually want.

HEAD-WORN DISPLAY:
A head-worn display (HWD) is a display worn on the head that uses the user's field of view to present digital content [Sutherland, 1968].

A bounded 2D display provides useful structure for window management.

These restrictions are not necessarily a drawback. A physical display bounds all windows to a small, fixed region and provides a stable reference frame. Its edges and corners serve as landmarks, and window positions repeat relative to this frame regardless of where the display is used. This spatial constancy lets users rely on spatial memory to find and switch between windows [Tak et al., 2009]. A 360° environment offers no such frame by default: Windows may sit anywhere around the user, including outside the current field of view, and layouts anchored to a specific room do not transfer between environments. The structure that spatial memory relies on must therefore be deliberately recreated. Bounded displays also support *physical navigation*, where users change their view by moving their body, for instance, by turning the head, leaning in for detail, or stepping closer. This navigation contrasts with virtual panning and zooming. On large displays, physical navigation has been shown to outperform virtual navigation [Ball et al., 2007], and the benefit stems from the movement itself rather than from the wider field of view a large display provides [Ball and North, 2008]. An HWD removes this bounded surface and, with it, the structure that spatial memory and physical navigation rely on. This removal of a bounded surface raises the question of which of these constraints are worth recreating in an open 360° space.

We focus on AR, where windows share the real environment.

Augmented Reality (AR) and Virtual Reality (VR) differ in how the surroundings appear to the user. In AR, the real environment remains visible and rendered objects share the space with the physical surroundings, whereas in VR the surroundings are replaced entirely by a computer-rendered environment [Milgram and Kishino, 1994]. For this work, we focus on the AR setting to uncover users' *mental model* of window and viewport management on an HWD.

AUGMENTED REALITY:
Augmented Reality (AR) combines virtual and real world content. This functionality can be achieved using HWDs, smartphones, and any device that can display digital content over a view of the real world. Interaction with the digital content happens in real time, and the virtual objects are rendered in three dimensions within the real environment [Azuma et al., 2001].

MENTAL MODELS:
Mental models are abstract models in people's minds of what they believe something is, how it works, and how to interact with it [Norman, 1983, 2002; Staggers and Norcio, 1993].

1.1 Window Management and Constraints

Constraints have shaped window management.

Constraints have always shaped window management. Historically, they came from computational limits. Early hardware and small screens could only support a full-window paradigm, in which a single window occupied the display at a time [Myers, 1988]. As devices became more capable and multiple windows could be shown at once, the limiting factor shifted to the display itself, which constrains window management to this day through its resolution, physical size, and, with multiple monitors, the bezels. HWDs can lift these physical limits, raising the question of which constraints are worth keeping.

Building on constraints, tiling-inspired approach.

Using Constraints for Window Management For this work, we build on several previous systems. On the desktop, Kandogan and Shneiderman's [1997] *Elastic Windows* organized windows in a space-filling, hierarchical tiling, where enlarging one window proportionally shrinks its neighbors and operations can be applied to whole groups. Closer to our setting, two studies used HWDs. Pavanatto et al. [2025a] compared a single boundary-free canvas against a multi-monitor layout on a *HoloLens 2* and found

that the canvas was clearly preferred in purely virtual setups. At the same time, monitor boundaries and snapping reduced arrangement effort. Pavanatto et al. [2025b] introduced *Spatial Bar*, a taskbar-like interface below a large virtual display that combines gaze-based selection with cursor teleportation to speed up switching between many open windows. Both studies constrained the workspace to a curved virtual monitor with a fixed depth, anchored to real-world coordinates and limited in width. Such bounded, user-centered layouts keep content within comfortable reach of head movement, avoiding the strain associated with wide layouts [Grubert et al., 2018].

TILING WINDOW MANAGEMENT:
A tiling window manager arranges windows in a way that they do not overlap. The available space is partitioned in a way to fit all the currently open applications onto the screen [Bly and Rosenberg, 1986; Cohen et al., 1986].

Recent AR studies focus on techniques and performance.

Most recently, two works have come closest to our setting. Luo et al. [2025] elicited interaction techniques for arranging documents in AR through a user-centered design study, cataloging how users would aggregate, distribute, and transform document collections. Marguet et al. [2025] built an Extended Reality (XR) window manager and compared interaction techniques for controlling it, focusing on system design and pointing performance. Both studies center on *how* users act on windows and documents, but neither examines what users expect a windowing system in AR to be in the first place. Pavanatto et al. [2025a] explicitly name the design of intelligent constraints for simplifying window placement on unbounded virtual displays as an open direction for future work.

Extracting the mental models of participants using a low-fidelity prototype.

In this work, we extract users' mental models to capture the windowing system they expect in AR, what they believe should happen, and how they intend to interact with it. Constraints inherited from physical displays cannot simply be carried over completely. While some provide structure worth recreating, others unnecessarily limit users, and some lose their meaning entirely once the physical screen

is gone. Eliciting users' mental models allows us to distinguish these cases and design a window management system around what users actually want, rather than around the legacy of the hardware.

Users accept structure they choose, not structure imposed on them.

Across both studies, one pattern proved central: Participants preferred free placement but did not reject structure as such. They rejected structure imposed on them. The formative study included a horizontally scrollable strip that fixed vertical position; participants found it restrictive precisely because it withheld space they could see was free, and they rated their sense of control under it far below free placement. However, the same participants welcomed structure they had chosen themselves, valuing it for the arrangement effort it saved. We therefore replaced the global band in our final prototype with user-controlled grouping and an optional, group-scoped scrolling mechanism. The distinction between structure that is offered and structure that is enforced runs through the remainder of this thesis.

1.2 Aim and Research Questions

With this thesis we aim to answer four research questions about viewport and window management.

Window management on the desktop and content placement in VR have been studied extensively. As discussed above, recent AR work has elicited interaction techniques [Luo et al., 2025] and evaluated system designs and pointing performance [Marguet et al., 2025]. However, users' *mental models* of window and viewport management on an HWD remain relatively unexplored. This thesis addresses this gap through four research questions.

Our formative study addresses the following exploratory questions:

RQ1: What mental model do users have of window and viewport management in AR?

RQ2: When arranging windows, how much and what kind of constraint do users want?

Building on the findings of the formative study, our second study evaluates concrete designs evolved from these findings:

RQ3: Do the constrained conditions improve the usability of window management compared to the free placement baseline?

RQ4: Which window management approach best matches users' mental models and preferences?

First, we elicit what users want; then, we implement and evaluate it.

We address these questions in two phases. To answer RQ1 and RQ2, we first conduct a low-fidelity elicitation and co-design study, focusing on the mental models users hold rather than the interaction techniques they use. From its findings, we derive three concrete window management designs: *Free Placement*, *Grouped Placement*, and *Scrollbar*. To answer RQ3 and RQ4, we then implement these designs in a high-fidelity prototype on the *Apple Vision Pro* and evaluate them in a second study.

This thesis offers four contributions.

From these two phases, this thesis makes four contributions:

- **It provides an empirical account of users' mental models of AR window management**, elicited on a low-fidelity prototype and condensed into five themes: Windows are sized and placed by importance, grouped by meaning with gaps as boundaries, retrieved from remembered locations, and understood through familiar desktop conventions, within a space users want free but structurable.
- **It develops the *SpatialWindowPlacer***, a native *visionOS* prototype realizing three window management designs on the *Apple Vision Pro*. Each design adds exactly one interaction mechanism over the previous one.
- **It evaluates these designs**, showing performance parity across all three conditions, no loss of perceived control under user-instantiated structure, a descriptive preference for opt-in grouping, and the rejection

of a scrolling mechanism that participants had themselves envisioned.

- **It offers two methodological observations for elicitation-driven design**: That low-fidelity elicitation predicted which *structures* users accept better than which *mechanisms* they adopt, and a technique in which each condition's placement rule is enforced by the physical material rather than by the experimenter.

The study materials, raw data, analysis scripts, and prototype source code for both studies are available online at this link[1].

Structure that is offered, not enforced.

One distinction recurs across both studies: What matters is not how much structure a system provides, but whether the user or the system decides on it.

1.3 Outline

In Chapter 2 *Related Work*, we review window management paradigms and prior HWD and AR/VR window-placement studies. Moreover, we introduce the foundations of mental models research. We also survey low-fidelity prototyping approaches.

In Chapter 3, we present the formative study, in which participants experience three different window management conditions on a low-fidelity prototype. Each condition is followed by semi-structured interviews, and the session closes with a co-design phase in which participants sketch their own preferred layouts. We report the questionnaire and interview results and close the chapter with the five themes that describe participants' mental models and a summary of findings with respect to RQ1 and RQ2.

In Chapter 4, we describe how these findings are translated into design decisions, and present the design and imple-

[1] https://hci.rwth-aachen.de/thesis/viewport-study/, *last accessed August 12, 2026*

mentation of the high-fidelity prototype on the *Apple Vision Pro*.

In Chapter 5, we evaluate the prototype in a second user study, report the quantitative and qualitative results, and discuss their implications for HWDs and AR window management research.

Lastly, in Chapter 6, we summarize our contributions, note the limitations of our studies, and propose directions for future work.

Chapter 2

Related Work

Our work draws on five areas of prior research. We first review how window management developed on desktop displays, where most of the conventions our participants bring with them were formed. We then follow the display beyond the desktop, to large and virtual displays, and to window and content management in AR. From this body of work, the question that drives this thesis emerges: How much and what kind of constraint do users want? Finally, we introduce the foundations of mental models research and the elicitation and prototyping methods on which our first study builds.

2.1 Window Management on Desktop Displays

Window management paradigms were shaped by hardware and display constraints.

Window management has always negotiated between limited display space and the user's need to work with multiple sources of information. Myers [1988] traced the development of window manager interfaces from systems that could show a single window at a time to overlapping and tiled approaches and provided a taxonomy of the design decisions involved. The two principal window layout paradigms of that era were tiled layouts, in which windows never overlap and the system partitions the available

space, and overlapping, in which windows float freely like paper on a desk. This distinction was empirically examined. Bly and Rosenberg [1986] compared both approaches and found that their relative merits depend on the task and on how much window arrangement overhead they impose on the user. User preference and task performance did not always favor the same paradigm. This early observation—that how people *like* to manage windows and how they *perform* with them can differ—resurfaces in the findings we present later.

Tiling can be automated through constraints.

Tiling lends itself to automation. Cohen et al. [1986] formulated tiled window layout as a constraint problem, letting the system maintain a valid arrangement automatically as windows appear and disappear. Kandogan and Shneiderman [1997] took a different route with *Elastic Windows*. Here windows are organized in a space-filling, hierarchical tiling in which enlarging one window proportionally shrinks its neighbors, and operations can be applied to whole groups of windows at once. Their evaluation showed advantages for multi-window operations over conventional window handling. Both systems show a recurring trade-off that motivates this thesis: Automatic structure reduces arrangement effort, but takes decisions away from the user.

Users often switch tasks; spatial consistency helps the user.

How people actually use windows is shaped by multitasking. Studies based on knowledge workers, whose work consists primarily of handling and producing information rather than physical output, show that it is fragmented into many concurrent activities. González and Mark [2004] observed that knowledge workers manage on average 10 different *working spheres* per day, and in a diary study, Czerwinski et al. [2004] documented how frequently tasks are switched and how difficult returning to an interrupted task is. Window switching is, therefore, a constant activity, and its efficiency depends on windows being where users remember them to be. Tak et al. [2009] showed that windows that keep stable, repeatable locations allow users to employ spatial memory when switching between windows. Stable positions let the workspace itself carry part of the burden: A window whose location is learned can be re-found by looking rather than by being held in mind. Where the

arrangement shifts unpredictably, that support falls away, and users must track positions internally instead, drawing on a working memory whose capacity is limited to a small number of chunks. The number of chunks is classically estimated at around seven [Miller, 1956], and in later work, it is closer to four [Cowan, 2001]. The cost of an unstable layout is, therefore, not merely the effort of rearranging it, but the effort of keeping track of it.

2.2 Beyond the Single Screen

Large displays work because users can physically navigate them.

One response to limited display space is more display. Research on large, wall-sized displays found that their benefit is not limited to the wider field of view. Ball et al. [2007] showed that promoting *physical navigation*, moving the body and turning the head, increases user performance on large-display tasks. Furthermore, Ball and North [2008] found that this physical navigation matters more than the added peripheral vision. Users, in other words, are good at using their bodies and spatial abilities to work with spatially distributed information, an ability that a 360° AR workspace can directly use.

Reference frames define where content is anchored.

Rather than building larger physical displays, HWDs render display space anywhere around the user. The idea itself is old; Feiner et al. [1993] placed 2D windows into AR three decades ago with *Windows on the World*, fixing them to the surrounding world, to the display, or to physical objects. We adopt this vocabulary throughout this thesis, using *world-fixed* for content anchored in the environment, *view-fixed* for content anchored to the head, and *body-fixed* for content anchored to the torso. View-management work uses *world-stabilized* and *screen-stabilized* [Bell et al., 2001].

Empirical layout studies favor world-fixed windows and fast switching.

Two decades later, Ens et al. [2014] evaluated such window arrangements empirically. Their *Personal Cockpit* arranges application windows on a curved surface within arm's reach, with each design parameter—window width, depth, angular separation, and spatial reference frame—tuned in a separate study. Contrary to their expectation that proprioception would favor body-anchored content, target

selection proved most precise in a world-fixed frame because body- and view-fixed windows drift with the reaching motion itself. The divergence Bly and Rosenberg [1986] observed on the desktop appears in AR as well: Targets placed below the start window were both slower and less accurate to reach; nevertheless, several participants preferred that direction because it reduced arm fatigue. Their final study had participants gather information distributed across several applications in a 3×3 or 4×4 window array, which the *Personal Cockpit* supported in roughly 60% of the time required by two view-fixed baselines, and 10 of 12 participants preferred it. This switching task is the paradigm our second study adapts (Section 5.1). Their studies ran in a CAVE emulating a 40° field of view rather than on an actual HWD, a limitation the authors name explicitly.

The virtual office is envisioned, but its organization is unresolved.

Grubert et al. [2018] sketched the modern version of this vision, a virtual portable office in which knowledge work is no longer bound to physical monitors. They simultaneously named its open problems, including the ergonomic strain of wide layouts and the question of how such workspaces should be organized.

On virtual displays, boundaries and snapping trade freedom against effort.

Two recent studies examine window management on such virtual displays directly, and both closely informed this thesis. Pavanatto et al. [2025a] compared a single, boundary-free canvas against a multi-monitor layout on a *HoloLens 2*. The canvas was clearly preferred in purely virtual setups, while monitor boundaries combined with a snapping feature reduced arrangement effort. The authors concluded that boundaries and snapping trade freedom against effort and highlighted the design of intelligent constraints for simplifying window placement on unbounded displays as a potential direction for future work. In another study, Pavanatto et al. [2025b] addressed window *switching* on a large curved virtual display. Their *Spatial Bar*, a taskbar-like interface below the display, combined gaze-based selection with cursor teleportation and proved fastest for switching over large distances. Both studies share a common workspace geometry, a curved, world-fixed virtual monitor of fixed depth and bounded width. This choice keeps content within comfortable reach of head movement and avoids the strain associated with wide layouts [Grubert

et al., 2018]. Thus, this workspace geometry keeps the display boundary from the desktop in AR rather than questioning it.

2.3 Window and Content Management in AR

View management decides where content may appear.

In AR, where virtual content shares space with the physical world, the placement question predates modern headsets. Bell et al. [2001] introduced *view management* for VR and AR, which maintains visual constraints on the projections of objects in the view plane, such as preventing labels from occluding each other. Their formulation treats placement as constraint satisfaction over regions of the view plane, an early precedent for the automatic approaches discussed later in this section.

Recent systems bring 2D windows and documents into immersive space.

Recent work brings full 2D windowing into immersive space. Marguet et al. [2025] built *WindowSpace*, a web-based XR window manager for arranging 2D application windows in immersive 3D space, and compared interaction techniques for controlling it, focusing on system design and pointing performance. Luo et al. [2025] approached the space from the content side. They conducted a user-centered design study in which participants proposed interaction techniques for arranging collections of documents in room-scale AR. Their participants defined groups through spatial proximity, treated layout compactness as meaningful, and proposed circular and even infinite stripe layouts navigated by scrolling. These findings converge with the mental models our participants expressed (Chapter 3).

Automatic layout optimization lacks user control.

A separate line of work automates AR layout entirely, placing content according to constraints such as environment geometry, semantics, or the user's perspective [Niyazov et al., 2023]. Niyazov et al. [2023] identified the central weakness of these approaches, namely the lack of user control over the optimization result, and proposed *user-driven constraints* as a remedy: Users specify constraints through

mid-air hand gestures directly in their physical environment, and the system re-optimizes content placement continuously as constraints are added. Their design space pairs eight constraints—attractive edge, repulsive edge, containment, exclusion, in-view, preference, user perspective, and semantics—with a region of interest (point, line, surface, or volume) and parameters such as weight, minimum distance, or capacity. In their summative study, all 12 participants preferred the constraint-based condition over manual placement, primarily for its perceived efficiency, and reported significantly lower temporal and physical demand without a statistically significant difference in perceived agency. Notably, all eight constraints relate content either to the physical environment or to the user's viewpoint; none of them express a relation between two content items. Alignment can still emerge as a by-product—windows attached to a shared attractive edge end up in a row, and participants in their study appreciated precisely this effect. Nevertheless, this alignment follows from anchoring each window independently to an environment feature rather than from any relation the model can represent. Constraints on the inner arrangement of the content itself, such as grouping, relative ordering, or layout grids, are explicitly left aside [Niyazov et al., 2023]. This thesis addresses this very gap.

Gaze needs an explicit confirmation, which motivates our input technique.

Beyond headsets, AR has also been studied on handheld devices, where the rendering perspective itself becomes a design decision. User-aware rendering couples the visible frustum to the user's head position, and the resulting field of view strongly affects how quickly users can search a virtual scene [Hueber et al., 2023]. Gaze tracking, in turn, has been used both implicitly, to place incoming content away from the user's current visual focus, and explicitly, to expand and dismiss it. Hueber et al. [2023] reported participants' reservations about gaze alone triggering actions. Their findings illustrates why gaze is typically paired with an explicit confirmation gesture, as in the gaze-and-pinch technique of *Apple Vision Pro* used in our second study.

2.4 Mental Models

Mental models can be incomplete and unstable, but they guide all interaction.

When users encounter a new system, they do not approach it without prior ideas. They bring a mental model of how it should work from previous experiences with other systems. Norman [1983] characterized mental models as naturally evolving internal representations that are incomplete, unstable, and partly superstitious. However, these are the models people actually use to predict a system's behavior. Staggers and Norcio [1993] surveyed the concept's use in human-computer interaction research, distinguished the user's mental model from the designer's conceptual model, and identified the system image through which users form their model. Norman [2002] popularized the design consequence; systems whose system image communicates a clear conceptual model allow users to form accurate mental models and are experienced as intuitive. For viewport and window management in AR, this framing raises the question this thesis poses: What model do users bring to a workspace that no longer has a physical screen?

Elicitation methods shape what they capture.

Mental models cannot be observed directly; they must be elicited, and the choice of elicitation method is consequential. Grenier and Dudzinska-Przesmitzki [2015] proposed a composite methodology because any single technique, such as interviews, drawings, or sorting tasks, captures only part of a model. Similarly, Doyle et al. [2022], in a review of elicitation methodologies across an entire field, concluded that methods shape and limit what is elicited. Two recognition-based techniques are particularly relevant to our study design. In card arrangement tasks, participants spatially arrange a set of cards until the layout matches their internal representation of a construct. This technique has been shown to effectively elicit internal knowledge structures across populations and domains, from children's understanding of technology [Marhan et al., 2012] to stakeholders' models of natural resource management [Jones et al., 2011]. The closely related diagramming approach instead asks participants to draw their understanding of a system and has been used, for example, to elicit users' mental models of the internet [Kang et al., 2015]. Our first study builds on both principles. Participants express workspace

layouts by physically arranging window cards and externalize their preferred designs through a co-design drawing phase. Following the composite reasoning above, we triangulate these data with interviews and questionnaires (Chapter 3). In AR specifically, elicitation has been used to gather interaction techniques [Luo et al., 2025] and gestures for defining layout constraints [Niyazov et al., 2023]. Our first study responds to the need to understand the underlying *mental models* of viewport and window management rather than interaction techniques.

Reflexive thematic analysis to turn elicited data into themes.

Turning the qualitative data into findings requires a systematic analysis method. We use reflexive thematic analysis [Braun and Clarke, 2006, 2019], which constructs themes as patterns of shared meaning across a data set and specifically treats the researcher's interpretative engagement as part of the method rather than a bias to eliminate [Braun et al., 2019]. The methodological details of our analysis are described in Chapter 3.

2.5 Low-Fidelity Prototyping

Low-fidelity prototypes trade realism for openness.

Our first study confronts participants with window management conditions before we develop that system, which is a classic use case for low-fidelity prototyping. Low-fidelity prototypes, such as paper or physical mock-ups, deliberately trade visual and functional realism for speed, cheapness, and openness. Because nothing about these prototypes looks finished, participants feel free to criticize, imagine, and propose alternatives. As a result, researchers can explore multiple design directions before committing to an implementation [Rettig, 1994; Snyder, 2003].

Physical prototypes have been used to study spatial arrangement before implementation.

For spatial arrangement questions in particular, physical materials have a second advantage: They make layouts tangible and manipulable without any tracking or rendering pipeline. Luo et al. [2025] used 50 physical paper documents for the context priming of their user-centered design study before eliciting techniques against predefined AR layouts. These virtual documents were deliberately view-only, without manipulation capabilities. Our first study

maintains the physical materials throughout, so that the elicited arrangements are performed on layouts that participants can actually rearrange with their own hands (Chapter 3). This choice trades AR fidelity for direct manipulability, and it accepts the legacy bias risk that Luo et al. [2025] name for their paper-based priming. We discuss this trade-off alongside the study in Section 3.4.

2.6 Summary

Prior work bounds the workspace. What users want remains open.

Taken together, the literature paints a consistent picture. Desktop research has established that window management negotiates freedom against arrangement effort [Bly and Rosenberg, 1986; Kandogan and Shneiderman, 1997] and that users rely on spatial constancy and physical navigation when working with spatially distributed content [Ball et al., 2007; Tak et al., 2009]. On HWDs, existing window management studies inherit the bounded, monitor-like workspace from the desktop, whether as a curved array within arm's reach [Ens et al., 2014] or as a virtual monitor of fixed depth and width [Pavanatto et al., 2025a,b]. Beyond this research direction, prior work compares interaction techniques for a workspace whose design is already settled [Marguet et al., 2025], elicits techniques for arranging content without asking what arrangement users want [Luo et al., 2025], or automates placement and restores control through user-defined constraints that stop short of the windows' inner arrangement [Niyazov et al., 2023]. None of these works establish what users themselves expect and want from window management once the physical screen is gone. What remains unknown is users' mental models of the space and the degree and type of constraint they would choose. The following chapters address this gap, first by eliciting those mental models with a low-fidelity study (Chapter 3), then by building the elicited structure into a working prototype (Chapter 4), and, finally, by evaluating it (Chapter 5).

Chapter 3

Eliciting Mental Models of AR Window Management

Before creating a layout, we had to find out what users want.

Viewport and window management is well established in 2D interfaces [Bly and Rosenberg, 1986; Kandogan and Shneiderman, 1997], and a growing body of work has explored it in AR and VR [Luo et al., 2025; Marguet et al., 2025; Pavanatto et al., 2025a,b]. These systems, however, largely present design solutions without establishing what users actually expect an AR layout to do. Designing a system around our own assumptions therefore risked working against the users' intuitions. To ground our design in evidence, we conducted a formative user study that elicited users' mental models of window management in AR across different conditions, using a low-fidelity prototype. This chapter presents the design and evaluation of that study. Section 3.1.1 describes the reasoning and decisions behind the study design. Section 3.2 presents the results, and Section 3.3 summarizes the findings with respect to both research questions. Chapter 4 describes how these findings shaped the *Apple Vision Pro* prototype. The study materials, collected data, and analysis scripts are available in the supplementary material[1].

[1] `https://hci.rwth-aachen.de/thesis/viewport-study/`

3.1 Methodology

This section describes the design and procedure we used for this formative study. It includes explanations of the different methods used and the conditions we tested in the study.

3.1.1 Study Goals and Approach

Prior work offers designs but not evidence of what users expect.

A review of prior work established how layout management has been approached in 2D interfaces and in AR/VR and where future work could be heading concerning layout management in AR (Chapter 2). Prior work offered design solutions, such as full XR window managers for placing 2D windows in immersive space [Marguet et al., 2025], but there is little evidence of what users expect such a layout to do. Our study addresses this gap.

We use four data sources because any single elicitation method is partial.

Our aim was therefore to understand how users expect an AR layout to behave, rather than to validate a design we had already chosen. To do so, we elicited participants' mental models (Definition 1). As any single elicitation method shapes and limits what it captures of a mental model [Doyle et al., 2022; Grenier and Dudzinska-Przesmitzki, 2015], we triangulated several data sources. Participants completed a questionnaire, a layout task with a low-fidelity prototype, semi-structured interviews, and a co-design drawing task. Two of these sources followed established recognition-based elicitation techniques (Section 2.4). The layout task builds on the principle of card arrangement, in which participants spatially arranged cards to express internal knowledge structures [Jones et al., 2011; Marhan et al., 2012], and the co-design task followed the diagramming approach, in which participants drew their mental model directly [Kang et al., 2015]. The individual measures are described in Section 3.1.6.

From these goals, we formulated two research questions:

RQ1: What mental model do users have of window and viewport management in AR?

RQ2: When arranging windows, how much and what kind of constraint do users want?

3.1.2 Study Design and Conditions

Prior work imposed constraints mainly for hardware and ergonomic reasons.

RQ2 asks how much constraint users want, so the central manipulation of the study is the degree of constraint imposed on the layout space. Prior work suggests that structuring the layout space can reduce the effort of managing windows. In 2D interfaces, constrained tiled layouts had faster performance than freely overlapping windows for some tasks and users [Bly and Rosenberg, 1986]. Structured multi-window layouts outperformed independent, freely arranged windows for a range of tasks [Kandogan and Shneiderman, 1997]. On large virtual displays, the presence or absence of boundaries measurably changes how users lay out and manage their windows [Pavanatto et al., 2025a]. We therefore examined what restrictions researchers have imposed in 2D and in AR/VR. Many were driven by hardware. In 2D, restrictions followed from computing power and the physical limits of the display [Myers, 1988], while in AR/VR they were driven by the need to mitigate neck and back strain [Grubert et al., 2018]. Pavanatto et al. [2025a,b] constrained the virtual workspace to a curved, monitor-like surface, limiting both its width and its height. We change this constraint in one respect. Rather than limiting the width, we allowed the workspace to extend laterally beyond the immediate view and to be scrolled, something our prototype approximated with its curved surface and movable table. To keep the low-fidelity prototype simple, window height was fixed in all conditions, and participants could vary only window width.

Three conditions as increasing degrees of constraint, each grounded in prior work.

From this, we derived three conditions that operationalize increasing degrees of constraint. As the unconstrained baseline, *Free Placement* mirrors the boundary-free canvas that participants preferred on virtual displays [Pavanatto et al., 2025a]. At the other end, *Fixed Horizontal* transfers the strict tiling paradigm of uniform, non-overlapping win-

dows [Cohen et al., 1986; Myers, 1988] onto the horizontal band. Between the two, *Elastic Horizontal* keeps the band but allows windows to vary in size, inspired by the proportionally sized tiles of *Elastic Windows* [Kandogan and Shneiderman, 1997].

C1 **Free Placement**: Participants could place windows anywhere in the workspace, at any of the three available sizes.

C2 **Fixed Horizontal** (internally: *Infinite Scroll*): Windows were constrained to a horizontal magnetic strip in the center of the prototype. The windows were always full size, and the workspace could scroll around the user.

C3 **Elastic Horizontal** (internally: *Elastic Scroll*): Windows were constrained to a horizontal magnetic strip in the center of the prototype. The windows had three different sizes, and the workspace could scroll around the user.

Condition names were chosen to avoid priming participants.

Participants were told only the neutral names *Free Placement*, *Fixed Horizontal*, and *Elastic Horizontal*. We deliberately avoided the internal names *Infinite Scroll* and *Elastic Scroll*, since naming the scrolling behavior could have primed participants toward it and biased the mental model we wanted to elicit. After all three conditions and their interviews were complete, we showed participants short videos illustrating how we envisioned each condition to behave. The videos can be seen here[2]. We discussed these in a follow-up interview (Section 3.1.5).

Study design and counterbalancing.

We designed the study as a within-subjects experiment [Lazar et al., 2017]. Every participant experienced all three conditions. To mitigate learning and fatigue effects, we used complete counterbalancing. We presented each of the 6 possible condition orders [Bradley, 1958]. For three conditions, this is equivalent to a balanced Latin square [Williams, 1949]. Orders were assigned cyclically

[2] https://www.youtube.com/playlist?list=PLbWON07TcMQU, *last accessed August 12, 2026*

by recruitment order, so that each of the six sequences was used by four participants. This step determined our sample size of $6 \times 4 = 24$ participants.

3.1.3 Apparatus: Designing the Low Fidelity Prototype

Low-fidelity prototype We wanted a prototype that was inexpensive and quick to build, while still allowing layouts to evolve. Other studies have presented participants with predefined, view-only layouts to elicit interaction ideas [Luo et al., 2025]. This approach would not let us observe how participants arrange windows themselves, or how a layout changes as windows accumulate. We, therefore, built a physical prototype that participants could manipulate directly.

Using acrylic glass to simulate the peripheral view.

The prototype consisted of three acrylic glass sheets, each 60 cm wide and 30 cm high, giving a total *screen size* of 180 cm×30 cm. We curved the sheets with a heat gun to resemble the curved virtual workspaces used by Pavanatto et al. [2025a,b], and mounted them upright on four 3D-printed stands. Placing the prototype on a table in the middle of the room preserved the participant's view of their surroundings, simulating an AR view rather than a VR view (Figure 3.1). The table could be moved, allowing participants to simulate the dragging and scrolling behavior of the workspace.

A curved acrylic surface on a movable table, keeping the room visible.

A magnetic strip was placed along the horizontal center of the acrylic surface. In the two constrained conditions, windows were attached to this strip magnetically, restricting them to the center of the view. Directly beneath the strip, a length of painter's tape substituted a scrollbar. We chose painter's tape because it could be applied and removed within seconds between conditions. The scrollbar was present only in the two constrained conditions, where the workspace could be scrolled, and it was absent in *Free Placement*, which offered nothing to scroll. We did not tell participants how this scrollbar behaved. Instead, we asked in the interview how they would expect it to work, so that

Magnetic strip enforces the constraint; scrollbar behavior left unexplained on purpose.



Figure 3.1: The finished prototype, positioned on a table in the middle of the room to simulate an Augmented Reality view, where a user can see the surroundings.

the expectation came from the participants rather than being imposed by us.

Participants need window representations that simulate the use of an HWD.

Window representations The layout task required windows that plausibly represented everyday computer use. Pavanatto et al. [2025b] used demo windows containing only numbers, which would not support realistic layout decisions. Luo et al. [2025] used a set of documents, which represents only a single type of window. Neither was sufficient for our purpose. To ground the window set in real usage rather than our own assumptions, we informally consulted with three colleagues at our lab, asking which applications they typically keep open during a working day. This measure was a lightweight step to select realistic windows, not a formal elicitation. Their input, together with empirical evidence that knowledge workers continu-

ally juggle multiple concurrent tasks and switch between them frequently [Czerwinski et al., 2004; González and Mark, 2004], informed a set of 20 windows. We grouped these into four categories of our own invention. The grouping serves to organize the windows for the study and does not follow a published taxonomy.

- Life Management: Contacts, Calendar, Mails, Banking Website, Home Assistant Application
- Productivity and Work: Word, Excel, Terminal, Slack, Zoom Website
- Information and Exploration: HCI Website, Maps, News App, Weather, Notes
- Leisure and Personal: YouTube, Music Player, Chess Game, Photos Application, Shopping Website

Creating window representations using Balsamiq.

We designed the window cards using Balsamiq[3]. To keep the representations close to real interfaces, we modeled them on the corresponding *macOS* applications and compared how each application renders at different window sizes rather than simply scaling a single image (Figure 3.2). Each of the 20 windows was created in three sizes, *Full Size*, *Two-Third Size*, and *Half Size*, giving 60 distinct window representations. The *Full Size* cards measured 16 cm×11.5 cm, with the two smaller sizes reduced in width (*Two-Third Size* 10.7 cm×11.5 cm, *Half Size* 8 cm×11.5 cm). All cards were laminated to withstand repeated placement, removal, and handling across the 24 sessions. We produced each card twice in identical form, for a total of 120 physical cards. One set had double-sided tape on the back, used in the *Free Placement* condition, where windows could be placed anywhere on the acrylic surface. The other set had a magnetic strip on the back, used in the two constrained conditions, where windows attach only to the magnetic strip of the prototype. Switching the card set between conditions thus enforced each condition's placement rule physically, without any instruction needed from the experimenter.

[3] https://balsamiq.cloud, *last accessed August 12, 2026*

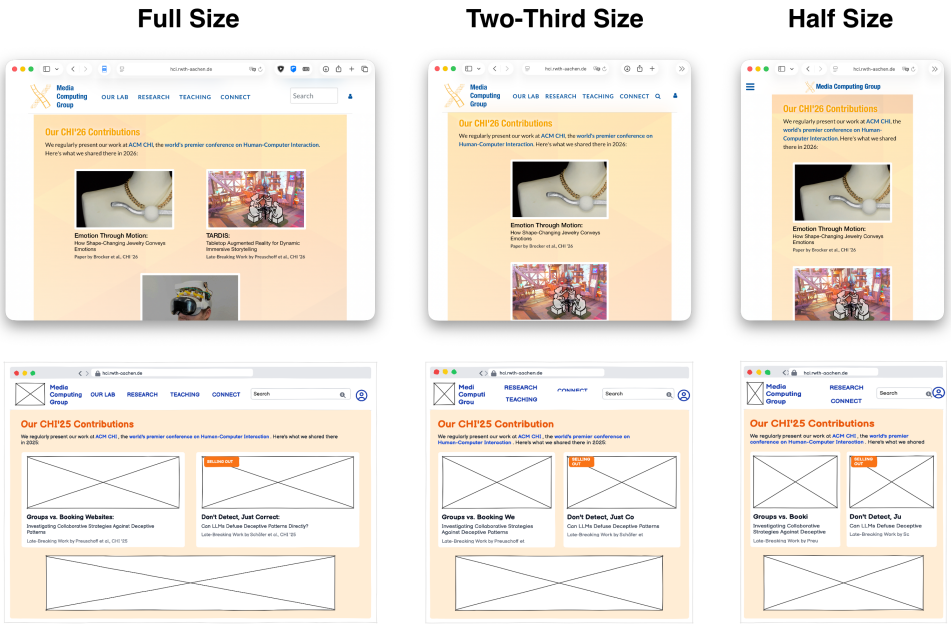


Figure 3.2: From application to window card: A website as rendered by *macOS* (top) and its recreation in Balsamiq for the prototype (bottom), each in the three sizes used in the study. Since applications reveal or hide interface elements depending on window size, each size was modeled on the corresponding *macOS* rendering rather than scaled from a single image.

3.1.4 Task and Scenario

We created a scenario to obtain a comparable baseline layout across participants.

Each condition consisted of two consecutive tasks. The first task was a scenario task, where participants planned a *weekend trip*. Rather than presenting all windows at once, we scripted a sequence of window events that participants carried out in order. This measure ensured that the workspace accumulated over time as it would during real work:

1. Open Maps
2. Open HCI Website
3. Open Calendar
4. Open Shopping Website
5. Open Mails

6. Open Weather
7. Open Zoom Website
8. *Close Shopping Website*
9. Open YouTube
10. Open Slack
11. Open Banking Website

Nine *Full Size* windows fit the strip, so overlap was a choice.

Ten windows were opened, and one was closed, leaving nine windows open at the end of the scenario. The window count also respected the physical capacity of the prototype. Even at the scenario's peak of nine simultaneously open windows, a gapless row of *Full Size* cards occupied 144 cm of the 180 cm strip, leaving room for two further cards. The constrained conditions, therefore, never forced participants to stack or overflow, and any overlap that occurred was a choice, not a necessity. Fixing both the scenario and the sequence made layouts comparable across participants, while the incremental order let us observe how a layout evolves as windows accumulate. The trade-off was naturalness: A scripted sequence constrained *which* windows appeared and *when*; consequently, we observed how participants placed windows they did not choose, not which windows they would have opened. The *free layout* task, which followed immediately within the same condition, was included to offset that problem. Here, participants arranged windows of their own choosing from the full set of 60 representations, with the condition's constraints still in force. Step 8 closed a window deliberately, letting us observe how participants handled removal from an existing layout.

Fixed Horizontal had a reduced window count because windows were only available in *Full Size*.

In *Fixed Horizontal*, this set was reduced to the 20 *Full Size* representations, as that condition permitted only full-size windows. The scenario ended with nine open windows, which nearly filled the strip in *Fixed Horizontal*. This allowed us to observe what participants do when space nearly runs out. Nine windows also exceeds common estimates of working memory capacity, whether the classic seven plus or minus two [Miller, 1956] or the lower estimate of approximately four chunks [Cowan, 2001]. This

was done so participants could not hold all open windows in mind at once and had to rely on the spatial layout itself.

3.1.5 Participants and Procedure

For the first study, we recruited 24 participants to elicit the mental model.

We recruited 24 participants through a direct approach and an announcement in our lab's Slack channel. Participants received refreshments during and after the session; no further compensation was provided. All participants gave written informed consent before the study, including consent to audio recordings. Participation was voluntary and could be aborted at any time. Data were pseudonymized; participants are identified only by number *(S1P1–S1P24)*, and no personally identifying information is reported. After transcription, the recordings were deleted. An ethics review was not necessary for this study based on the guidelines of the FÜEK Ethics Committee, as it involved no vulnerable groups, no deception, and no more than minimal risk. Participant demographics are reported in Section 3.2.1.

The session took place in a room at our lab. The room measured 3.65 m×4.90 m×2.75 m (width × length × height). We recorded the interviews using an *iPhone 12 Pro*. Participants completed the questionnaires themselves on an *M1 MacBook Pro* using a locally hosted web form, and photographs were taken with an *iPhone 13*.

Procedure: Interviews first, then videos to preserve mental model elicitation.

Each session lasted approximately 90 minutes and followed the same sequence. Participants first read an information sheet and gave written informed consent. They then completed a short questionnaire capturing demographics, monitor setup, prior VR/AR experience, and a free-text description of their everyday multitasking habits, and they were briefed on the prototype and the layout task. Participants completed all three conditions in their assigned order. After each condition, a semi-structured interview explored the layout they had created and their reasoning behind it. Once all three conditions were complete, a further interview asked participants to compare the conditions. We then showed three short videos illustrating how each condition

was envisioned to behave in a working system, followed by a brief interview capturing participants' reactions. Since these videos revealed the intended scrolling behavior, they were shown only after all conditions and the participant interviews had concluded (Section 3.1.2). Finally, participants completed a co-design drawing task, sketching their own preferred AR layout, and a short interview about the drawing. The session closed with a preference ranking where participants ranked the three conditions from most to least preferred, once overall and once for each of three work contexts (writing and research work, creative and design work, and communication and email).

3.1.6 Measures and Analysis

We gathered several data points to triangulate our findings.

As Section 3.1.1 states, any single elicitation method shapes and limits what it captures of a mental model [Doyle et al., 2022; Grenier and Dudzinska-Przesmitzki, 2015]. We therefore collected four complementary data sources for triangulation:

- **Questionnaires**: A pre-session questionnaire captured demographics (age, gender, occupation), number of monitors used, prior VR/AR experience, and a free-text description of everyday multitasking habits. A post-session questionnaire was also administered in which participants rated each condition on eight items using a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree): suitability for writing and research, for creative design, and for communication and email; how quickly the layout was understood; its predictability; the sense of being in control; how mentally demanding it was; and its overall usefulness. Participants then ranked the three conditions overall and separately for each of the three task types. The questionnaires can be seen in Appendix A.
- **Interviews**: Semi-structured interviews were conducted after each condition, a comparative interview after all three, a short interview following the concept videos, and an interview about the drawing.

- **Drawings**: A co-design drawing task was conducted in which participants sketched their preferred AR layout.
- **Photographs**: The layout created in each condition was photographed.

The first three data sources were used to elicit participants' mental models and were triangulated during analysis. The photographs were collected as documentation and serve illustration purposes only.

Interviews were analyzed using reflexive thematic analysis.

Interviews Interviews were audio recorded and transcribed with noScribe[4] and checked by the author of this thesis. Transcripts were analyzed inductively using reflexive thematic analysis following Braun and Clarke [2019], using the six phases originally described by Braun and Clarke [2006] in an earlier work. The answers were open-coded by the author using Taguette[5], yielding 77 codes across 2532 coded segments; the full codebook is given in Appendix B. Coding was data-driven. Consistent with reflexive TA, we treated themes as patterns of shared meaning organized around a central concept rather than as domain summaries and judged their significance by analytic meaning rather than by frequency [Braun et al., 2019]. Coding proceeded across all 24 participants.

Photographs were not coded because they did not always match the user's mental model.

Photographs We photographed the end-of-task layouts after each condition. These were not coded because a photograph does not convey intent independently of the interview: It can show a layout that differs from the one the participant had in mind. One participant, for example, explained that a visible gap in their layout existed only because they had not moved the windows together, and that in their mind no gap was present.

> *"No, this, this, like this is supposed to be not a gap. Ups. So, uh, my main screen, there's usually no*

[4] https://noscribe.de/en/, *last accessed August 12, 2026*
[5] https://www.taguette.org, *last accessed August 12, 2026*

> *gap. This gap is for me, I'm imagining this as other screens"*
>
> —S1P24

Analytic claims about layouts, therefore, always rested on the interviews, with photographs illustrating what participants described.

Drawings were coded separately and source-tagged to preserve data origin.

Drawings Each drawing was transcribed into a structured written description of its observable content, recorded consistently across participants. Only what was visible was described; interpretations of intent were excluded, and anything a participant explained verbally was captured in the interview data instead. These descriptions were then coded using the codes developed from the interviews, supplemented by codes specific to the visual data (for example, the drawn grouping device used in each layout). Every drawing-derived code was additionally tagged with a source marker, so that highlights originating in drawings remained distinguishable from those originating in interviews. Groupings across drawings emerged from this coding rather than being imposed in advance.

Questionnaires were Friedman and Wilcoxon tested with Bonferroni correction.

Questionnaire All analyses were carried out in a *Python* script. Each of the eight rating items was compared across the three conditions using a Friedman test, appropriate for repeated ordinal (Likert) measures. To guard against Type-I errors across the eight items, we applied a Bonferroni correction ($\alpha = .05/8 = .00625$). Where a Friedman test remained significant after this correction, we ran Wilcoxon signed-rank post-hoc tests on the three pairwise comparisons, Bonferroni-corrected them ($\alpha = .05/3 = .0167$), and reported the rank-biserial correlation r as an effect size. Because tied absolute differences occurred in every pairwise comparison, the Wilcoxon tests were evaluated using the normal approximation with a tie correction rather than the exact distribution. For the omnibus tests, we reported Kendall's W. Condition rankings were analyzed the same way, converting each participant's ranks to scores *(first = 3, last = 1)* and correcting across the four ranking categories

(α = .05/4 = .0125). Finally, to check that the presentation order had not biased ratings, we ran the same Friedman procedure on ratings grouped by the position *(first, second, third)* at which a condition was experienced.

3.2 Results and Evaluation

3.2.1 Demographics

24 participants, most describing a central primary task with secondary content at the periphery.

Table 3.3 summarizes the demographics of our sample. Among the 23 participants with prior VR or AR experience, the most common were gaming and *Apple Vision Pro* use (each reported by 17 participants), followed by HoloLens (3) and other devices (4); several participants reported more than one device. When asked to describe their multitasking habits in an open question, 23 of the 24 participants responded. A consistent pattern emerged: Most participants described a single primary task placed centrally or on a main monitor, with secondary content such as communication, music, or reference material arranged at the periphery. Often, the secondary content was on additional screens or devices. Several participants reported keeping windows in fixed, role-assigned positions:

> *"My windows always have a "fixed" position, e.g. I always open slack on the right and the browser on the left."*
>
> —S1P8

Participants	24
Age	M=25.4, SD=3.0 (range 22–32)
Gender	13 male (54%), 11 female (46%)
Immersive experience	23 some (96%), 1 none (4%)
Number of monitors	1: 5 (21%), 2: 14 (58%), 3+: 3 (13%), varies: 2 (8%)

Table 3.3: Demographics of our first study with the low-fidelity prototype.

3.2.2 Quantitative Results

Order effects No item showed an effect of presentation position, either before or after correction (all $p > .11$), so ratings are reported independently of order.

Free Placement rated highest on 4/8 items with the largest effect in creative work and control.

Condition ratings Four of the eight items differed significantly across conditions after Bonferroni correction, all favoring *Free Placement* (Table 3.4). Figure 3.5 shows the full response distributions per item. The largest effects were for creative and design work ($\chi^2(2) = 32.4$, $p < .001$, $W = .68$), overall usefulness ($\chi^2(2) = 25.9$, $p < .001$, $W = .54$), and sense of control ($\chi^2(2) = 25.6$, $p < .001$, $W = .53$). Writing and research showed a moderate effect ($\chi^2(2) = 18.8$, $p < .001$, $W = .39$). Post-hoc tests largely followed the same ordering, namely *Free Placement* being the highest, *Elastic Horizontal* second, and *Fixed Horizontal* the lowest, with the free–fixed contrast reaching a near-maximal effect for creative work and control ($r > .95$). Two pairwise contrasts did not survive correction: *Free Placement* against *Elastic Horizontal* for writing and research ($p = .029$) and the two constrained conditions for sense of control ($p = .084$). Two items showed no significant difference: Suitability for communication and email ($p = .218$) and mental demand ($p = .646$). Two further items, how quickly the layout was understood and its predictability, were significant only before correction and did not survive it.

Rankings The rankings reinforced the ratings (Figure 3.6). *Free Placement* was ranked first overall by 20 of 24 participants ($\chi^2(2) = 30.1$, $p < .001$, $W = .63$), and by all 24 for creative and design tasks ($W = .86$). *Fixed Horizontal* was ranked last overall by 20 of 24 participants, and last in 78 of the 96 rankings across the four categories. *Elastic Horizontal* occupied the middle position in every category: significantly preferred to *Fixed Horizontal* throughout, and significantly below *Free Placement* in all but communication and email ($p = .106$).

	Free Pl.		Fixed Hor.		Elastic Hor.	
	Mdn	*M* (±*SD*)	*Mdn*	*M* (±*SD*)	*Mdn*	*M* (±*SD*)
Writing/Research *	5.0	4.29 (±1.00)	3.0	2.96 (±1.04)	4.0	3.62 (±0.77)
Creative/Design *	5.0	4.67 (±0.56)	2.0	2.25 (±0.94)	3.0	2.96 (±1.12)
Communication	4.0	4.12 (±0.99)	4.0	3.75 (±0.90)	4.0	4.00 (±0.66)
Quickly Understood	5.0	4.83 (±0.38)	5.0	4.42 (±0.78)	4.0	4.21 (±0.93)
Predictability	4.0	3.88 (±0.90)	3.0	3.25 (±1.03)	3.0	3.38 (±1.01)
Felt in Control *	5.0	4.67 (±0.64)	2.0	2.83 (±1.24)	3.5	3.21 (±1.28)
Mentally Demanding	3.0	2.75 (±0.99)	3.5	2.96 (±1.16)	3.0	2.83 (±1.01)
Overall Useful *	4.5	4.33 (±0.76)	2.0	2.58 (±1.14)	3.0	3.21 (±0.83)

Table 3.4: Medians, means, and standard deviations for the eight post-session questionnaire items per condition, rated on a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree). * marks items with a significant Friedman test after Bonferroni correction ($\alpha = .00625$). Note that for *Mentally Demanding*, lower values are the favorable outcome.

3.2.3 Qualitative Results

Creating five themes with TA, triangulated the interviews with drawings and questionnaire.

Reflexive thematic analysis of the interviews constructed five themes describing participants' mental models of AR viewport and window management (Table 3.7). We report each in turn, illustrated with representative participant extracts and triangulated against the layout drawings and the questionnaire results. A sixth set of observations, concerning limitations of the prototype itself, is reported separately in Section 3.2.4, as it reflects on our study's low-fidelity prototype rather than on AR layouts. One participant declined the drawing task, stating that the *Free Placement* condition already matched their preferred layout; therefore, the analysis includes 23 instead of 24 drawings. Some interviews were conducted in German. Quotes from these were machine translated using DeepL[6] and subsequently verified by us. Translated quotes are marked with a *Translated* marker next to the participant number.

[6] https://www.deepl.com/, *last accessed August 12, 2026*

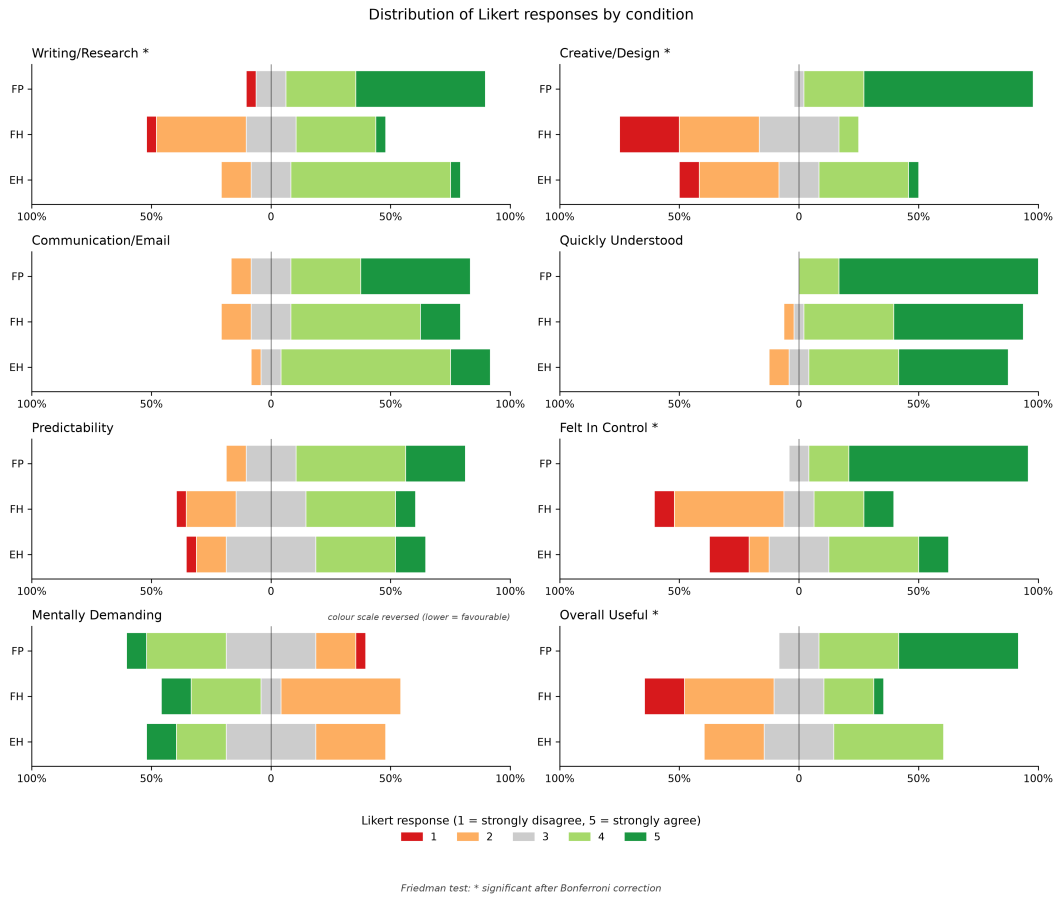


Figure 3.5: Distribution of the Likert responses for the eight post-session items per condition (FP = *Free Placement*, FH = *Fixed Horizontal*, EH = *Elastic Horizontal*). With $N = 24$ participants. Bars diverge around the neutral midpoint. * marks a significant Friedman test after Bonferroni correction ($\alpha = .00625$). Note that for *Mentally Demanding*, a lower value is a favorable outcome.

T1: Importance Influences Size, Place, and Visibility

Across the interviews, participants consistently tied a window's importance to how they treated it spatially. Importance was not an abstract label but something expressed through three concrete decisions, as follows:

- How large the window was,
- Where the window was placed, and

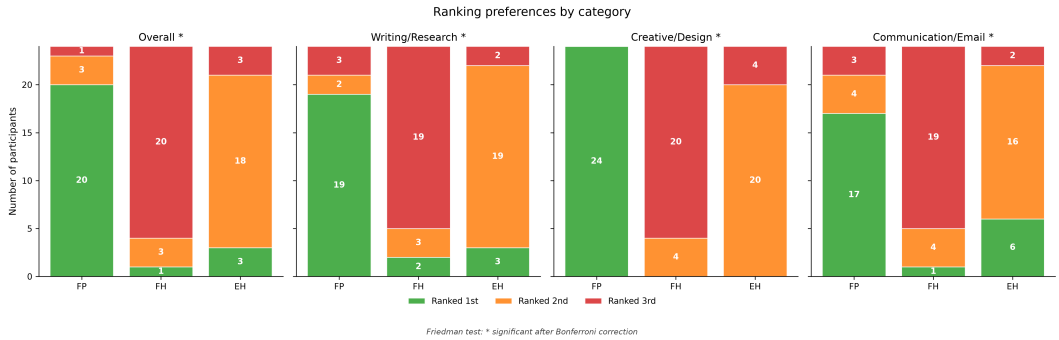


Figure 3.6: Ranking preferences per task category (N = 24 participants). Stacked bars show how often each condition (FP = *Free Placement*, FH = *Fixed Horizontal*, EH = *Elastic Horizontal*) was ranked first, second, and third. * marks a significant Friedman test after Bonferroni correction for four ranking categories (α = .0125).

- Whether the window remained in view throughout the task.

The first theme indicates that importance dictates size, placement, and visibility of windows.

The clearest expression was size. Participants scaled windows to their importance, enlarging what they were working on and shrinking what they only needed to glance at. As S1P6 described it:

> *"So the things that have my main focus will always be the largest, while some things, where I don't need to interact as much, I can just quickly glance to."*
>
> —S1P6

The reasoning was explicitly one of hierarchy. A window's size showed its priority, so that the most important content could be seen in full, while peripheral information was smaller, kept to the minimum needed to read it.

Unneeded windows were moved aside, not closed.

Importance also influenced visibility. Windows that were no longer needed were not necessarily closed but moved out of the way while remaining recoverable. S1P18 explained that they would "like, either minimize it to some kind of application bar or just, like, put it behind me somewhere" (S1P18). Removal from the central view was thus treated as an action rather than an open or closed state, with less important windows pushed to the periphery view.

T1	**Importance influences size, place and visibility:** A window's importance governs how large it is, where it sits, and whether it stays visible.
T2	**Window Organization:** Participants impose structure on the space, grouping related windows and using gaps and enclosures as boundaries.
T3	**Bringing Windows Into View:** How participants move content into view, via head movement, scrolling, search, and moving windows to focus.
T4	**Using Desktop Paradigms:** Participants fall back on familiar 2D desktop logic and metaphors to make sense of AR.
T5	**Freedom vs Structure:** The experience swings between the freedom of open space and feeling restricted or overwhelmed.

Table 3.7: The five themes constructed through reflexive thematic analysis of the interview data, each capturing a pattern in how participants understood and managed windows in AR.

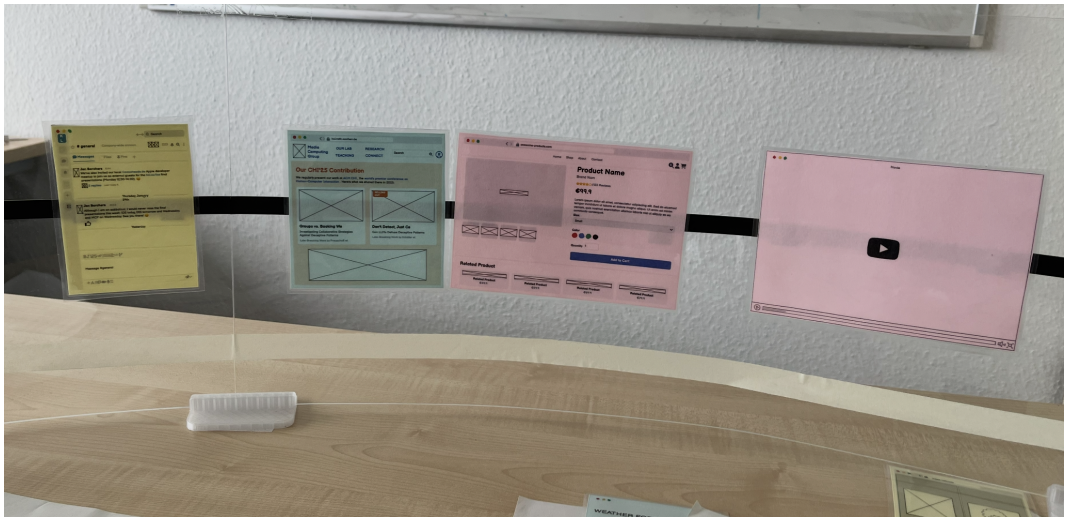


Figure 3.8: A participant's layout in the free layout task of the *Elastic Horizontal* condition. Windows were attached to the strip in all three sizes, illustrating the size differences participants used to express a window's importance (T1).

Size and central placement together mark the current focus.

Finally, importance determined placement. The most important window was repeatedly described as belonging directly in front of the user, in the center of their view. S1P24 described it as follows:

> *"So my main applications are always big and in front of me. so I can directly see them."*
>
> —S1P24

Here, size and placement reinforced each other. The central position and the large size together marked a window as the current focus.

Free Placement rated highest, consistent with importance-driven layout.

Taken together, these accounts describe a mental model in which *importance* is the organizing principle behind the layout. It is read off the size of a window, its position relative to the user, and its presence in or absence from the field of view. This reading is reflected in the questionnaire results, where *Free Placement*, the only condition that allowed participants to freely size and position windows by importance, was rated highest for both control and overall usefulness (Section 3.2.2).

The drawings showed a similar pattern. A single central or dominant window appeared in 17 of the 23 drawings (74%). Furthermore, 21 (91%) participants used visibly different window sizes rather than a uniform grid, consistent with participants' sizing and placing windows by importance.

T2: Window Organization

The second theme reveals thematic grouping, deliberate gaps, and context-dependent overlap.

While T1 focused on individual windows in a layout, T2 focuses on the relationships between them. Faced with open space, participants did not scatter windows arbitrarily but imposed structure on them. They clustered related windows, separating them with deliberate gaps, and layering them through overlap.

The most common structuring strategy was thematic grouping. Participants placed related windows together,

organizing the workspace by the meaning of its contents rather than by any spatial rule. S1P1 described grouping windows by function:

> *"Yeah, so most of the stuff were, like, thematically together. So, if I have, like, a messenger and, like, email, it's, like, communication stuff. It goes into one section."*
>
> —S1P1

Given freedom, participants grouped by meaning rather than by spatial rule.

This semantic organization was something several participants reached for once given freedom to arrange the space themselves. S1P11 noted:

> *"And also, once you gave me everything and said I could change it, then I was also happy to group them more semantically."*
>
> —S1P11

Gaps marked boundaries between groups.

The deliberate use of gaps reinforced grouping. Rather than packing windows together, participants left spaces between them to mark boundaries. For S1P14, the gap was what created a group:

> *"I organize my groups just by having space between them."*
>
> —S1P14

Most of the time, gaps were not empty leftover space but an active structuring device, used to signal where one cluster stopped and another began.

This ordering was also reflected in the drawings. Twenty-one of 23 participants grouped their windows (91%) and 22 of 23 left gaps or separation between them (96%). Moreover, 11 of 23 drawings (48%) marked groups with an explicitly drawn enclosure such as a box or loop.

A third strategy was overlapping windows, which was more prominent in the interviews than in the drawings



Figure 3.9: A participant's layout at the end of the scenario task in the *Free Placement* condition. Windows were arranged in two rows above and below the strip and clustered thematically, with the communication windows placed together on the right and gaps separating the clusters (T2).

Overlap was a desktop habit but not welcome under constraint.

(below). Several participants treated overlap as a natural, space-saving habit carried over from desktop use. S1P20 explained:

> *"I tend to like to overlap things as a habit from having limited space [...] I'd rather just have things overlap a tiny bit and still see a bit more of the window behind it. So that when I need it, I can just click on it and it comes to the front."*
>
> —S1P20

This preference was not universal and not preferred under certain constraints. In the *Fixed Horizontal* condition, where windows were confined to the central strip, some participants rejected overlap outright, with S1P22 stating:

> *"I don't see myself liking overlaps here."*
>
> —S1P22

Overlap appeared in the interviews but rarely in the drawings.

This interview preference for overlap was not seen in the drawings, where only six of the 23 layouts (26%) depicted

any overlap at all. The remaining 17 (74%) were drawn without showing any overlap. We read this divergence not as a contradiction but as informative. The two data sources, thus, describe the same participants from different angles, namely how they describe managing a workspace and how they picture an idealized one.

Taken together, these accounts show participants treating the open workspace as something to be actively ordered. Grouping and gaps were often tools for imposing an order, while overlap, though widely used, remained sensitive to context.

Same grouping principle recovered in room-scale AR, not just an artifact of our apparatus.

The same grouping logic appears in Luo et al.'s [2025] elicitation study, where participants also defined groups through spatial proximity and treated layout compactness as meaningful. Their participants stood in a room wearing a *HoloLens 2*, surrounded by 110 holographic documents they could not touch; ours sat at a table moving laminated cards by hand. That two studies differing this much in fidelity, scale, and modality recover the same organizing principle suggests that proximity-based semantic grouping belongs to users' models of spatial layout rather than to either apparatus.

T3: Bringing Windows into View

The third theme revealed three retrieval mechanisms. Carousel metaphor common, but some motion-sickness concern.

Once a workspace holds more windows than it fits comfortably in front of the user, some interaction is needed to bring a window into view. Participants described three such mechanisms, differing in what actually moves: The system repositions windows, the user moves windows or the workspace, or the user moves themself.

Automatic movement welcome for periphery, but must never disturb the focal window.

The first was automatic movement, where the system rearranges windows on the user's behalf. Participants imagined that new windows opened into the center and existing ones yielded space and that gaps closed themselves when a window was removed. S1P7 described both, where a new window would open in the middle:

> *"So when I open a new window, I would, um, imagine that it would open in the middle, maybe like here or here. And what it would do is that the things get pushed kind of out of the way or a bit, a little bit to the side."*
>
> —S1P7

Similarly, while closing a window, a participant said:

> *"Uh, so how I would imagine to change is that this closes, it goes away and then the gap would kind of automatically close as well. Uh, so everything would automatically shift so that the gap would close."*
>
> —S1P7

However, automatic movement was not completely accepted and had clear limits. It was welcome only where it did not disturb the current focus. S1P9, who kept a near-full-screen game in front of them, wanted new windows placed aside rather than centrally:

> *"So, I would hate that anything would open directly in front of it [. . .] And if there's space available, then open it somewhere there in the free space."*
>
> —S1P9

Automation was, in general, desired for tidying the periphery, but not when it moved the focal window.

Scrolling imagined as carousel rotation and some wanted per-window control.

The second mechanism was scrolling, understood by most participants as moving the whole workspace at once rather than any single window. The scroll mechanism was only present in the two constrained conditions. Some participants also included a scrolling mechanism in their drawings. They consistently imagined a single gesture that shifted everything together, and reached for circular, egocentric metaphors to describe it. S1P15 pictured the workspace as revolving around them:

> *"I would like to, basically, everything around me to spin. Like, all of the things remain at the same point in the circle, like, in a 360 degree circle, and just this moves to in front of me"*
>
> —S1P15

S1P9 used the same image more briefly, imagining the layout behaving "[...] just like a carousel" (S1P9), where all windows move identically. The recurring notion was not scrolling a single window but rotating one's surrounding windows. A smaller number of participants wanted finer control, moving an individual window rather than the whole space. S1P9 added, "I would also love to have some control over the individual windows" (S1P9), and S1P21 imagined that a single window moves at the same speed as their gesture:

> *"And if I just move a window, it just moves the window on the same trajectory."*
>
> —S1P21

Physical turning as retrieval; spatial memory and personal responsibility.

The third mechanism did not move windows at all. Participants moved their body or head to find what they wanted. We asked how they would locate an off-screen window. Some participants, like S1P4, answered, "I think I would just turn like my body" (S1P4). For S1P15, the answer was tied to spatial memory and even a sense of personal responsibility for the layout:

> *"Looking around. Because I still try to keep, like, the most important stuff in my peripheral sideline. It's hopefully as easy as turning around once. If I, again, I said it, if I have put it at, like, random places, I probably had a reason. And if not, then I deserve to have more effort finding it."*
>
> —S1P15

Retrieval, in this view, was a matter of remembering where one had placed something and physically turning to it rather than summoning it.

The three mechanisms differ in what moves, not in what users want.

Across all three mechanisms, the underlying concern was the same, namely keeping the currently relevant window reachable without losing track of the rest of the windows. Participants differed in whether they wanted the system, the windows, or themselves to do the moving. One common sentiment was the expectation that the workspace should make the necessary window easy to bring to their focus.

Motion-sickness concern which informed group-scoped design.

The interviews following the concept videos (Section 3.1.5) added a caveat to the scrolling mechanism. While most participants confirmed that the envisioned behavior matched what they had imagined, a small number voiced concern about the entire workspace moving around them, anticipating that it could cause discomfort:

> *"I think scrolling would make me motion sick, maybe"*
>
> —S1P20

This concern did not surface for movements of individual windows, only for motion of the surroundings as a whole, and it informed our decision to scope scrolling to groups rather than the workspace in the high-fidelity prototype (Section 4.1).

Carousel metaphor also appears without a strip in another study, so our apparatus did not supply it.

Participants in Luo et al.'s [2025] study independently proposed circular and stripe-like layouts navigated by scrolling, despite working in room-scale AR with no strip, no scrollbar, and no bounded surface to suggest one. Our participants reached for the same carousel metaphor while facing a 180 cm curved sheet. The convergence is notable because our apparatus could plausibly have produced the image on its own; that it also arises without any such affordance suggests the egocentric, rotating workspace is a model users brought in rather than one our prototype supplied.

T4: Using Desktop Paradigms

A recurring feature of participants' mental models was that they were not built from scratch for AR. Instead, participants reached for the conventions of the desktop and monitor setups they already knew. Here, they mapped familiar two-dimensional strategies onto the three-dimensional space.

Participants often mapped their own personal monitor setups.

This transfer was often explicit. Participants described arranging the AR workspace exactly as they would arrange their screens at home. S1P10 stated:

> *"I'd say I did it here almost exactly like at home, the way I'd do it on my PC with a multi-monitor setup. With a single-monitor setup I'd probably tend to minimize everything and pull the windows back up one at a time from the minimizer."*
>
> —S1P10 Translated

Notably, the reference point was not a *desktop* in the abstract but each participant's own specific setup. Participant S1P10's behavior differed depending on whether they imagined their multi-monitor or single-monitor habits.

Borrowed patterns were concrete mechanisms, like switching virtual desktops.

The paradigms borrowed were often concrete interaction patterns rather than vague resemblances. S1P7, working in the constrained *Fixed Horizontal* condition, mapped it onto a familiar *macOS* mechanism:

> *"It would be comparable to if I, um, used just my, uh, MacBook and have, like, multiple, like, full screen windows open, I would just, like, cycle through them, like, to the next virtual screen."*
>
> —S1P7

Here an AR interaction was understood by direct analogy to switching between virtual desktops.

Familiar reliance was a conscious decision.

For some participants, this reliance on the familiar was a

conscious principle rather than a habit. S1P17 described how old strategies persist in new technologies and are adapted:

> *"I'm a big fan of using old strategies with new technologies. And then eventually you'll get a new strat for the new tech. But, like, generally if a new tech can also do the stuff that the old stuff could, then it's not gonna, like, establish itself."*
>
> —S1P17

On this account, a new medium must first support what the old one did, and new interactions must follow later.

Mental models anchored in prior desktop experience.

Taken together, these accounts suggest that participants' mental models of AR window management were substantially anchored in prior desktop experience. This finding has a direct implication for our design. Rather than treating users as approaching AR without preconceptions, a system can expect them to arrive with established two-dimensional habits. We will return to this point in the design of the *Apple Vision Pro* prototype (Section 4.1).

A desktop-style element, *taskbar*, *dock*, or *menu bar* was also present in 13 of the 23 drawings (57%). This element was another sign that participants imported familiar two-dimensional conventions into the AR space.

T5: Freedom vs Structure

Fifth theme revealed that freedom is preferred but structure welcomed if opt-in.

The final theme concerns how participants experienced the open workspace itself. Their accounts describe how the freedom of a large, unconstrained space was preferred, but that too much freedom can sometimes become a burden. Rather than preferring either extreme, most participants wanted freedom that could be bounded.

Open space valued for personal control and tailoring.

The freedom of the open space was, in general, preferred. Participants valued being able to place and size windows however they liked, not only using it to place things but also arranging them more to their liking.

> *"I think a lot, actually, because I have way more freedom as to, like, I could choose way more freely where to distribute the single windows in which sizes. So, I have way more, let's say I have way more options to tailor this experience to my liking."*
>
> —S1P6

The participants experienced the open space as positive. More room meant more control over how the workspace fit the individual participants.

Constrained conditions could feel obstructive when they withheld visibly available space. S1P12 found the central horizontal band frustrating because unused space sat just out of bounds:

> *"Yeah, I feel like the horizontal line was too restrictive. Just, I have all the space over here, why do I, like, if, like, considering everything here is full and I need another important window, I would have to move over here, well, there's just space right there."*
>
> —S1P12

Here, restriction was experienced as an arbitrary hindrance rather than a helpful structure.

Constraint welcomed when it reduced effort.

However, restriction was not always disliked. This finding shows the theme's central tension. Several participants valued a degree of constraint, not as a limitation but as a source of structure that reduced effort. S1P2 reflected that being restricted could be productive:

> *"I think in this specific scenario, I felt this restricted setup. But maybe not in a bad way. Because if you are restricted, maybe you follow more rules for yourself."*
>
> —S1P2

Sometimes, the preference was not to have total freedom:

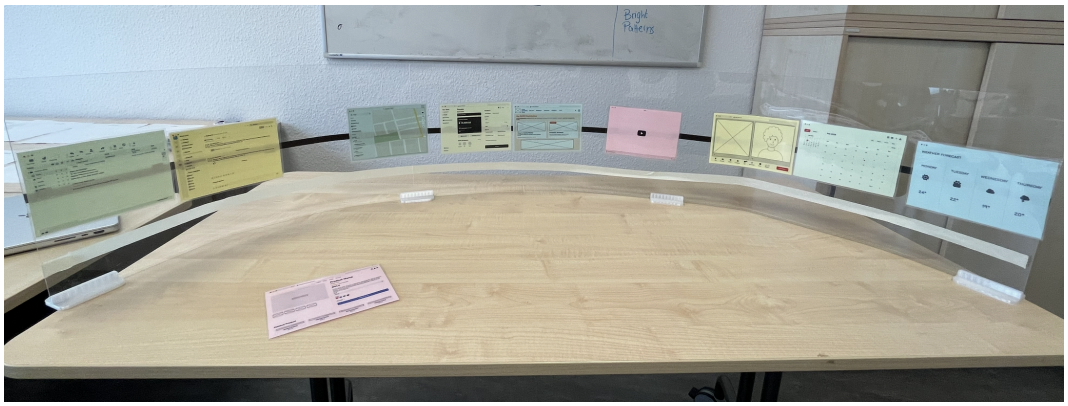


Figure 3.10: A participant's layout in the *Fixed Horizontal* condition. Full-size windows filled the magnetic strip and the scrollbar under it while the space above and below remained unusable, illustrating the restriction participants experienced as withholding visibly available space (T5).

> *"I like the concept of free placement, but maybe not too much free placement."*
>
> —S1P24

Enough structure to save effort but enough freedom to tailor.

This finding suggests that we should provide enough freedom to tailor the space and express oneself, but enough structure to avoid the effort of arranging everything by hand. The participant's assertion directly informs our second research question on the kind and degree of constraint users want (Section 3.1.1). Participants did not always want an unconstrained space but a bounded one. Therefore, we designed the *Apple Vision Pro* prototype with this middle ground in mind (Section 4.1).

Figure 3.11 shows four participant drawings that illustrate the recurring patterns described above. The drawings are not a representative sample but were chosen to make the themes visually accessible. A full-screen view of these drawings is also provided in the Appendix C.

3.2.4 Prototype Limitations

The prototype was not perfect; some participants noted differences to their mental models.

Other than just mental models, participants raised issues concerning the low-fidelity prototype itself. We report

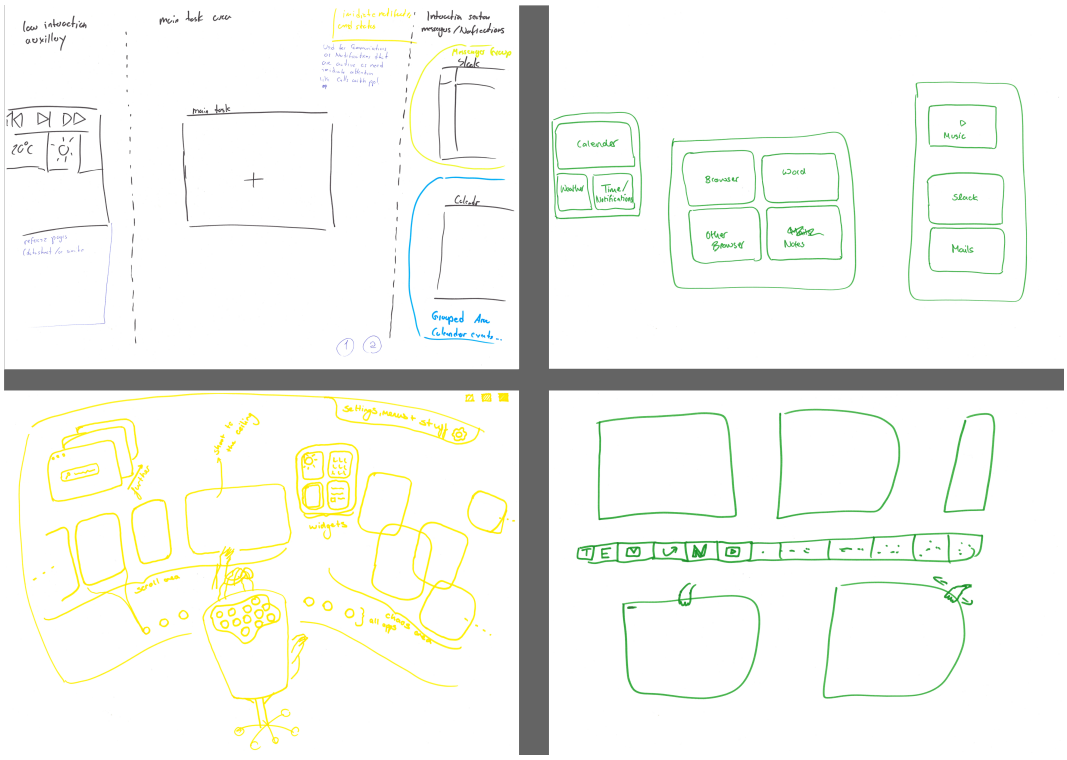


Figure 3.11: Participant drawings from the co-design task illustrating four of the five main themes. (Upper-Left) A dominant central window (T1). (Upper-Right) Explicit group boxes and separating gaps (T2). (Bottom-Left) A circular layout navigated by turning or scrolling (T3). (Bottom-Right) A taskbar-like element in the workspace (T4).

these concerns separately because they reflect on the study apparatus rather than on window management in AR. They also informed the design of the *Apple Vision Pro* prototype (Section 4.1) and are considered again in the limitations of this study (Section 3.4).

The given window sizes of the prototype were not dynamic enough.

The issues participants raised concentrate on a single aspect. The three discrete window sizes could not express the layouts they had in mind. Most prominently, the largest available size was not large enough. Several participants wanted to enlarge individual windows well beyond it:

> *"If I could make chess bigger, I would."*
>
> —S1P17

> *"I would like to have an even larger option for a window. So like wider, maybe even taller."*
>
> —S1P23

The fixed window height was a further constraint. Finally, one issue concerned the fidelity of the representations rather than their size. For S1P20, the proportions of the *Notes* window did not match the application they knew, which made the card irritating to place.

The size complaints strengthen T1 and mark a known limit of the next prototype.

Two things follow from this feedback. First, it strengthens rather than weakens T1. Participants did not merely use size to express importance within the offered range; they pushed against the range itself, wanting size differences that were more extreme than the prototype allowed. Second, it marks a boundary for the high-fidelity prototype. Window sizing there remains limited to three sizes as a scope decision (Section 4.1), so this feedback is carried forward as a known limitation rather than a solved problem. Thus, free continuous resizing remains an avenue for future research.

3.3 Summary of Findings

What the study answered, and what it means for the next chapter.

The formative study set out to answer two research questions (Section 3.1.1). The five themes constructed from the interviews, triangulated with the drawings and questionnaire results, allowed us to answer both questions. After that, we discussed the limitations of this chapter.

RQ1 answered by T1-T4. Importance-based, semantic, and anchored in desktop conventions.

RQ1 asked what mental model users have of window and viewport management in AR. The answer was captured by the first four themes, which together form a coherent picture rather than four separate observations. Participants organized a layout around importance. A window's size, its position, and its presence in or absence from the field of view all expressed how much it mattered at that point to a user (T1). They structured the space semantically, clustering related windows and using gaps as boundaries (T2). To

bring content into view, they expected one of three mechanisms, differing in what moves: These were that the system rearranges windows, the user moves windows, or the whole workspace, or the user turns toward a remembered location of a window (T3). These expectations were not built from scratch for AR but anchored in familiar desktop and multi-monitor conventions (T4).

RQ2 answered by T5. Opt-in structure is preferred because imposed constraints reduce control.

RQ2 asked how much and what kind of constraint users want when arranging windows. The answer lies in T5 and the quantitative results. Participants preferred freedom: *Free Placement* was rated highest on four of eight items and ranked first by 20 of 24 participants (Section 3.2.2). However, the interviews showed that this preference was not for unbounded space. Constraint was rejected when it withheld visibly available space but welcomed when it provided structure that reduced the effort of arranging everything by hand. The constraint users wanted was, therefore, opt-in structure within a free space, or bounded freedom, rather than either extreme. The same tension recurs in systems research. Users of virtual displays preferred a boundary-free canvas for its freedom, yet valued the structure of monitor boundaries and snapping for reducing arrangement effort [Pavanatto et al., 2025a]. Niyazov et al. [2023] offered a complementary datapoint: When users themselves defined constraints, all participants preferred the constraint-supported arrangement over manual placement for its efficiency, with no significant loss of perceived agency. The comparison is indirect: Their constraints attracted content toward regions the user chose, whereas our *Fixed Horizontal* forbade an axis of placement outright. Therefore, the two studies differ both in who defines the constraint and in how restrictive it is, and the measures differ as well. Taken together with our results, where the imposed constraint of *Fixed Horizontal* sharply reduced the sense of control, the pattern nonetheless suggests that structure per se does not cost users their sense of control, but imposed structure does. That an elicitation study with a physical prototype, an evaluation of a working AR system, and a constraint authoring system converge on the same pattern suggests it is robust rather than an artifact of any single method.

3.4 Limitations

The findings reported above should be interpreted within the limits of the study's method, sample, and material.

Prototype captures layout expectations, not interaction performance.

Fidelity of the prototype The most important limitation is the gap between the physical low-fidelity prototype and a real AR system. Our prototype did not provide a depth axis, as is available in an AR system. We only provided a fixed window height with three discrete window sizes that users could use. Windows were moved by the users by hand rather than by gaze and gesture or by using some kind of controller like in other AR/VR HWDs. The scrolling behavior only existed as an imagined behavior and was simulated by moving the table on which the prototype stood. We argue that this gap is acceptable for this study, but it is not neutral. Mental-model elicitation targets participants' expectations about how a workspace should be *organized*. We wanted to identify users' expectations regarding window placement, grouping, and how content should return to view. These expectations can be elicited using physical cards, as in our study, as well as using holograms, as demonstrated by Luo et al. [2025]. What our prototype cannot inform are claims about *effort, precision, comfort, or performance*. Manipulating a card says nothing about manipulating a window through pinch gestures, and we accordingly make no such claims in this chapter. The practical issues our participants raised with the apparatus (Section 3.2.4) highlight this boundary, and the high-fidelity study in Chapter 4 addresses the interaction side that the cards could not. Another consequence of the missing headset is that participants always saw the entire 180 cm workspace at once and had no full 360° workspace like in an HWD. A real HWD's limited field of view would hide parts of it and could change both layout behavior and the perceived need for retrieval mechanisms (T3).

Legacy bias is not removed. The conditions frame what is elicited.

Elicitation shapes what is elicited Any elicitation method shapes and limits what it captures of a mental model [Doyle et al., 2022; Grenier and Dudzinska-

Przesmitzki, 2015]. We mitigated this issue by triangulating multiple data sources, but two design choices deserve mention. First, we deliberately did not counteract *legacy bias*, the influence of prior interfaces on what participants propose. Whether such bias should be reduced has been debated. Morris et al. [2014] proposed techniques to counteract the bias, which some technique-elicitation studies in AR have adopted [Luo et al., 2025]. In contrast, Köpsel and Bubalo [2015] argued that designs benefit from it, as familiar patterns ease the transition to new forms of interaction. Our reason is a third one. Since our goal was to capture the mental models users bring with them, including their desktop conventions, this transfer is the object of study rather than a confound, and the desktop paradigm borrowing in T4 is, accordingly, a finding. The counterpoint is that our results describe the expectations of today's desktop-trained users and may shift as spatial computing becomes more familiar. Second, the three conditions themselves frame what participants could react to. Expectations about constraints not represented in any condition, such as depth layering or automatic layout movement, only surfaced if participants volunteered them.

Young, VR-experienced sample whose results reflect early adopters.

Sample Our 24 participants were rather young (aged 22–32 years), predominantly from a university context, and almost all had prior VR/AR experience. These participants are representative of early adopters of AR, but they do not represent the broader population of knowledge workers. Older users or those without multi-monitor experience may hold different models, and T4 in particular may look different for them. The sample size was chosen for complete counterbalancing rather than statistical power, and the quantitative results should be read as convergent evidence alongside the interviews, not as a stand-alone survey.

Single scenario domain where findings may be task-dependent.

Task and materials The scenario task covered a single domain, *planning a weekend trip*, with a fixed sequence of 10 windows. This window set was curated by us, informed by an informal consultation rather than a formal elicitation (Section 3.1.3). The questionnaire itself suggested task dependence. Suitability ratings differed across writing, cre-

ative, and communication tasks (Section 3.2.2), so layouts for other domains may follow different rules. The free task partially compensated for this difference by allowing participants to choose their own windows but within the same set of 20 applications.

Themes are one interpretation, supported by triangulation and external evidence.

Analysis Reflexive thematic analysis treats the researcher as an instrument. The themes are our construction from the data, not the only possible reading [Braun and Clarke, 2019; Braun et al., 2019]. We supported the analysis with source-tagged triangulation across interviews, drawings, and questionnaires, which strengthened the credibility of the resulting themes. In addition, our T5 aligns with independent systems research by Pavanatto et al. [2025a], providing further confidence in these findings.

These limitations qualify the extent to which the findings generalize, but they do not undermine the study's purpose. Our goal was not to measure performance in AR but to ground the design of an AR window management system in users' expectations. The next chapter opens by translating the five themes into concrete design decisions (Section 4.1) and then describes the implementation of the prototype in detail. Limitations that concern the thesis as a whole, including those shared by both studies, are discussed in Chapter 6.

Chapter 4

The High-Fidelity Prototype

The formative study in Chapter 3 served not only to describe mental models but also to ground the design of our *Apple Vision Pro* prototype in the evidence of what users actually expect. We implement this grounding in this chapter. The prototype source code is included in the supplementary material[1].

4.1 From Findings to Design

How the low-fidelity prototype study shaped the high-fidelity prototype

The elicitation of participants' mental models produced five themes (Table 3.7):

- *T1* Importance expressed through size and placement
- *T2* An open space actively ordered through grouping and gaps
- *T3* Three retrieval mechanisms, differing in what moves: the system, the windows, or the user
- *T4* Familiar desktop conventions carried into AR

[1] `https://hci.rwth-aachen.de/thesis/viewport-study/`

T5 A consistent preference for bounded freedom over either extreme

This section translates these themes into design decisions. We first describe how and why the prototype departs from the conditions of the low-fidelity study. Then, we discuss how the themes shaped the prototype's overall structure and its three conditions.

4.1.1 From Horizontal Constraint to Free Space

For the mental model, we wanted to use a strict constraint of a horizontal band. With our findings, we decided against restricting it this strictly.

The low-fidelity study constrained two of its conditions to a horizontal band in the center of the workspace. The findings argued against carrying this idea forward. Participants consistently preferred *Free Placement*, and the band constraint was experienced as withholding visibly available space (T5). The structure participants valued was not a global band but local, semantic organization. Through the interview and drawings, we determined that participants preferred to have related windows grouped, separated by deliberate gaps (T2). Although some layout drawings contained horizontal band-like arrangements, the accompanying interviews indicated that these reflected drawing convenience rather than a desired constraint. We therefore abandoned the idea of a global band, which we will still discuss in the future work chapter (Chapter 6). The prototype gave users a free, unbounded workspace in every condition. We reintroduced structure not as a property of the workspace but as something users create within it, through groups. This functionality gave users the freedom to choose how restricted they wanted to be.

4.1.2 Conditions as a Response to Freedom and Structure (T5)

The freedom-structure tension became the prototype's organizing principle.

The clearest signal from the study was T5. Most of the time, participants wanted total freedom but also the possibility of structure. We made this tension the organizing principle of the prototype by implementing it in the form of three

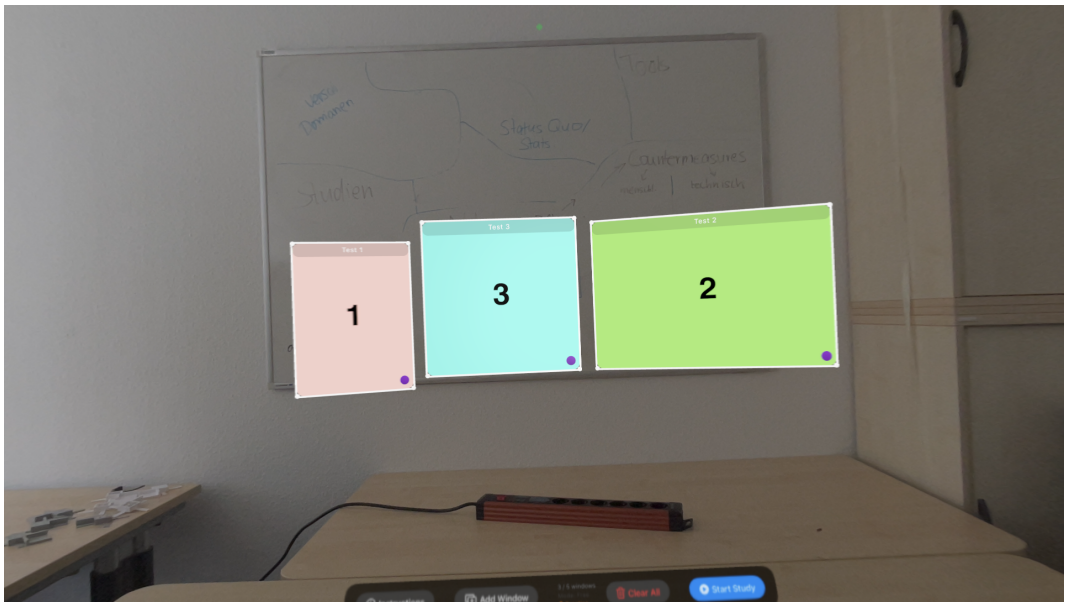


Figure 4.1: The *Free Placement* condition during the practice phase, showing the three window sizes (*Half Size*, *Two-Third Size*, and *Full Size* at similar height). Windows could be placed anywhere in the space. The purple button on each window cycled its size starting from *Full Size*, to *Two-Third Size*, to *Half Size*. Five colored, numbered windows were the practice set used before the proper study.

conditions that progress along a freedom-to-structure approach, which allowed us to focus our attention on the middle ground participants described.

We developed three different conditions for the *Apple Vision Pro* prototype

The first condition, *Free Placement*, offered an unbounded arrangement and approximated the *Apple Vision Pro's* standard interaction. In this condition, windows could be placed and sized anywhere (Figure 4.1). The second condition, *Grouped Placement*, retained this freedom but added the ability to create structure. Here, windows could be snapped into groups and taken out of them again, letting users constrain windows inside a bounding box if they so choose (Figure 4.2). The third condition, *Scrollbar*, extended *Grouped Placement* with a scrolling mechanism for shifting windows within a group (Figure 4.3). Each condition, therefore, preserved the freedom participants valued while offering an increasing degree of structure which they also asked for.

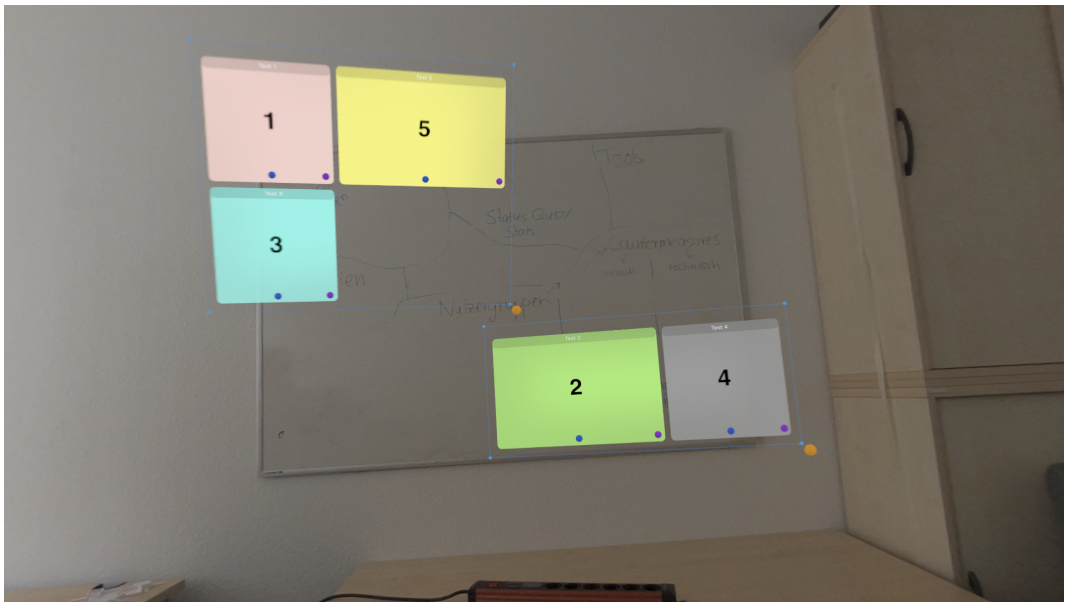


Figure 4.2: The *Grouped Placement* condition with two groups of different arrangements. Each group was enclosed by a bounding box and organized as a grid that could contain empty cells. The orange button at the box corner compacted the group by removing empty rows and columns that were left by other windows. Each window carried a blue detach button and a purple size button.

4.1.3 Window Behavior (T1)

Users control size, position, and visibility; sizing stays limited to three steps.

T1 showed that participants read a window's importance from its size, position, and presence in their view. The prototype reflected this by giving users direct control over all three aspects. A window opened at *Full Size* and directly in front of the viewer; the position and scale created the current focus for the participant. After that, the windows could then be freely repositioned and resized. To let participants size windows, each window was available in three sizes of differing width, cycled through using a per-window button. This set of three was also a deliberate scope decision. It reused the window representations created for the low-fidelity study. Moreover, it kept the implementation easy within the scope of this thesis while still supporting the importance-driven sizing identified in T1. The size limitation participants criticized in the low-fidelity prototype (Section 3.2.4), therefore, persisted here as a known constraint. However, the constraint was

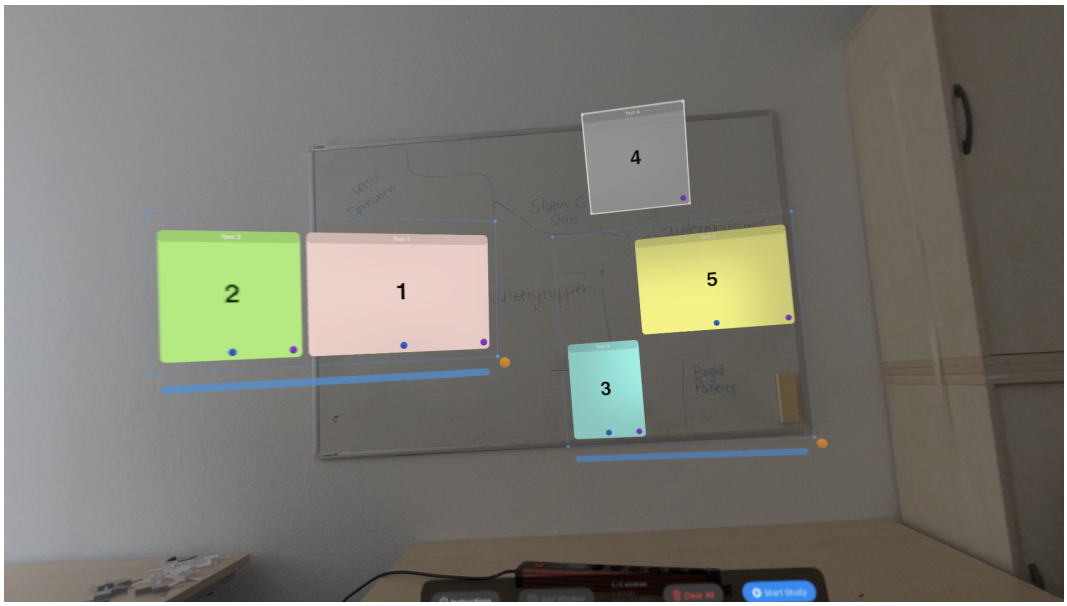


Figure 4.3: The *Scrollbar* condition. Each group additionally showed a blue scrollbar below its bounding box. Dragging the scrollbar shifted the group's windows left or right, with windows wrapping around at the ends like a carousel. Ungrouped windows (top) behaved as in *Free Placement*.

slightly mitigated by the possibility of also changing the Z-axis, which was not possible in the paper prototype.

4.1.4 Grouping and Overlap (T2)

Snapping creates groups; empty cells preserve gaps, closed only on request.

T2 established that participants actively organized the workspace by grouping related windows and separating them with deliberate gaps. In the *Grouped Placement* and *Scrollbar* conditions, dragging a window close to another window or an existing group allowed the user to snap it in. Before the snap was committed, a red preview line indicated the edge at which the window would be inserted before the user committed to this choice (Figure 4.4). A bounding box enclosed groups and could be manipulated as a single unit. The groups' internal structure was a grid that participants extended freely in both directions. Group size was unbounded so that users could decide their own group sizes. Grids could contain empty cells, preserving the gaps participants used as meaningful separators (T2).



Figure 4.4: Snapping and overlap handling. The dragged window (Test 2, green) was rendered translucent so the window behind it stayed visible. A red preview line on the right edge of the stationary window indicated where the dragged window would snap into place when released.

Closing these gaps was never automatic but triggered with a compact button on the group itself.

Overlap permitted; translucent dragging added after the pilot.

The drawings and interviews of the formative study diverged on overlapping. Participants sketched tidy, non-overlapping layouts, but described overlapping windows as natural. Overlap was, therefore, permitted, following S1P20's wish to partially overlap windows and "[...] still see a bit more of the window behind it" (S1P20). A window additionally became translucent while it was being dragged (Figure 4.4). This decision came out of the pilot session, where we found that a dragged window hid both the content beneath it and the red snap preview line (Section 4.2).

4.1.5 Bringing Windows into View (T3)

Scrolling scoped to groups which preserves retrieval by turning.

T3 described three mechanisms by which participants expected to bring content into view, and two of them shaped

the prototype directly. Participants imagined scrolling as rotating content "just like a carousel" (S1P9). The *Scrollbar* condition implemented exactly this behavior. Dragging the bar below a group rotated the group's windows stepwise through their reading order, with a short glide animating each step of the rotation. We deliberately scoped scrolling to the *inside* of groups rather than the whole workspace, for three reasons. First, when reacting to the concept videos in the formative study, a small number of participants voiced concerns that the entire surroundings shifting around them could induce cybersickness. One participant stated, "I think scrolling would make me motion sick, maybe" (S1P20), a concern consistent with the known role of large-field visual motion in inducing such discomfort [LaViola, 2000]. Second, restricting scrolling to groups kept the three conditions minimally different; each condition added exactly one mechanism over the previous one. In this way, differences between adjacent conditions remained attributable to that mechanism. Third, a workspace-wide scrolling mechanism would have exceeded the implementation scope of this thesis. Whether scrolling the entire workspace around the user, a *globe* of windows, is viable without discomfort remains an open question we return to in the future work discussion in Chapter 6. Furthermore, participants expected to find windows by turning toward their remembered location. The free, world-anchored workspace preserved this, and the platform's gaze-and-pinch interaction was used for all selection and manipulation.

CYBERSICKNESS:
Cybersickness describes symptoms of discomfort, such as nausea, disorientation, and eye strain, that can occur during the use of virtual environments. It is a form of visually induced motion sickness. The visual system perceives self-motion, for example when the surroundings move around the user, while the body remains stationary. Unlike classic motion sickness, no physical movement is required to trigger it [LaViola, 2000].

T3 also surfaced a split over automatic movement. Some participants expected the system to close gaps and shift

Gap-closing made with a button press rather than automatic.

windows on its own, while others rejected any movement they had not initiated. We, therefore, made gap-closing a choice-based rather than automatic feature. When a window was removed from a group, its cell remained empty until the user pressed the group's compact button, which removed empty rows and columns. This method followed S1P21's idea:

> *"It's fine if I have a button to close the gap. It's not fine if it closes the gap on its own"*
>
> —S1P21

The only automatic adjustments the system performed were small corrections. A newly formed group shifted slightly to sit fully in view, and scroll steps glided briefly rather than teleporting.

4.1.6 Desktop Familiarity (T4)

Groups act like virtual desktops, a taskbar drawn by over half of participants.

T4 showed that participants approached AR through familiar desktop conventions, and the prototype leaned on this in two ways. Groups functioned as virtual zones, echoing the virtual desktop and multi-monitor arrangements participants referenced. In addition, the prototype provided a taskbar (Figure 4.5), the desktop element that appeared in over half of the co-design drawings (13 of 23). In previous research, a taskbar-like element has also proved effective for window access on large virtual displays [Pavanatto et al., 2025b]. In the prototype, the taskbar appeared at a fixed offset below the user's view and carried the controls for opening windows as well as the study controls described below. Users could also freely reposition the taskbar if they wished.

4.2 Implementation

The prototype, *SpatialWindowPlacer*, is a native *visionOS* application targeting *visionOS 26.5* built with *SwiftUI*, *Reali-*

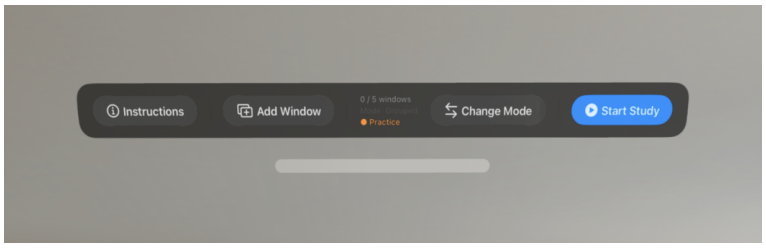


Figure 4.5: The taskbar, positioned in front of and below the user's view. Left to right: Reopening the current instructions, adding the next window, a status readout (windows placed, current mode, practice/study phase), switching the mode, and Start Study.

visionOS prototype *SpatialWindowPlacer* using immersive space.

tyKit, and *ARKit*, running on a first-generation *Apple Vision Pro*. All windows live in a single immersive space as textured entities rather than native OS windows, which gives the application full control over placement, snapping, and grouping. The application logs all interactions unobtrusively in the background. Head position is tracked through ARKit's world-tracking provider and used to spawn windows facing the user and to position the taskbar. Selection and dragging of elements use the platform's standard gaze-and-pinch input [Pfeuffer et al., 2017].

Reusing the low-fidelity study windows and sizes. Comparability maintained between the studies.

Window Representations The prototype reused the 20 windows of the low-fidelity study as screenshot textures, edited for legibility on the headset and removing the *macOS*-style taskbar. A hotel-search website replaced the HCI website from the low-fidelity study to fit the scenario. Each window existed in the same three sizes as the physical cards: *Full Size* (0.6 m), *Two-Third Size* (0.4 m), and *Half Size* (0.3 m) at a constant height of 0.4 m. Windows were oriented toward the user when placed and retained that orientation until they were moved again. A per-window button cycled the sizes. Like all interactive controls in the prototype, the button tried to emulate the conventions of native *visionOS* elements. This feature was not just cosmetic. With gaze-and-pinch input, the effective width of a target was limited by eye-tracking accuracy, and studies of gaze-supported selection in Fitts' law tasks accordingly used and

recommended generously sized targets that absorb gaze endpoint scatter [Wagner et al., 2023; Wei et al., 2023]. Keeping the same window set with one substitution and the same size scheme tied the two studies together. Layouts, sizes, and search targets were, therefore, directly comparable across fidelity levels. For the practice phase that preceded each session, five nondescript, colored test windows were used instead (visible in Figures 4.1, 4.3), so that participants could try the mechanics without previewing the study windows.

Creating the different interaction methods for the prototype.

Snapping and Groups Snapping was proximity-based. When a dragged window came within 15 cm of a window or group edge, a red preview line appeared at the insertion position. After releasing the window, a snap was committed, and the group's grid was rearranged with small uniform gaps between members. Groups maintained their members in an unbounded $N \times M$ grid. Columns remembered their width so that detaching and re-snapping windows did not reshuffle the remaining layout, and empty cells persisted until compacted. A group could be dragged as a whole by its bounding box. Detaching used the per-window detach button; a group was dissolved implicitly when the last two members separated. Windows inside a group shared the group's orientation rather than each turning toward the user, so that the group presented a single flat plane.

Creating a smooth moving scroll behavior for grouped windows.

Scrollbar In the *Scrollbar* condition, each group displayed a horizontal bar slightly below its bounding box. Dragging the bar rotated the group's windows through their reading order in discrete steps. Each dragging step was animated with a short glide (0.18 s) so that windows visibly moved to their new cells rather than teleporting. Scrolling was deliberately scoped to the group (Section 4.1).

Application drives the study and logs all interactions.

Study Support and Logging Beyond the interaction mechanics, the Application embedded the machinery for the evaluation in Chapter 5. The taskbar carried the controls

(mode selection, starting the study, phase progression) and showed participants their current instruction. The scripted *weekend-trip* scenario and the *search task* were driven by the Application itself. During search trials, every window displayed a green search button, and participants answered a trial by tapping the button on the target window (Figure 4.6). Taps on any window were logged with their correctness, which created the selection accuracy and search-time measures. Every interaction was logged as a timestamped event stream (one JSONL file per participant and condition), including:

- Placements and drags
- Snap previews and commits
- Group formation and dissolution
- Detachments
- Compact-button presses
- Scroll steps
- Size changes
- Search-trial timing
- Final session layout

These logs served as the basis for the measures reported in Chapter 5.

The pilot study refined the prototype, changing transparency and gliding animation.

Pilot and Refinements Before the study, we conducted a pilot session with one member of our lab, who did not participate in the evaluation study, and refined the prototype based on their feedback. Two changes resulted. First, we made windows translucent while being dragged. In the pilot, the opaque window hid both the content behind it and the red preview line, making it hard to judge where a window would snap. Second, we changed the scroll behavior. Initially, scrolling teleported windows to their new positions, which produced a flashing effect when performed

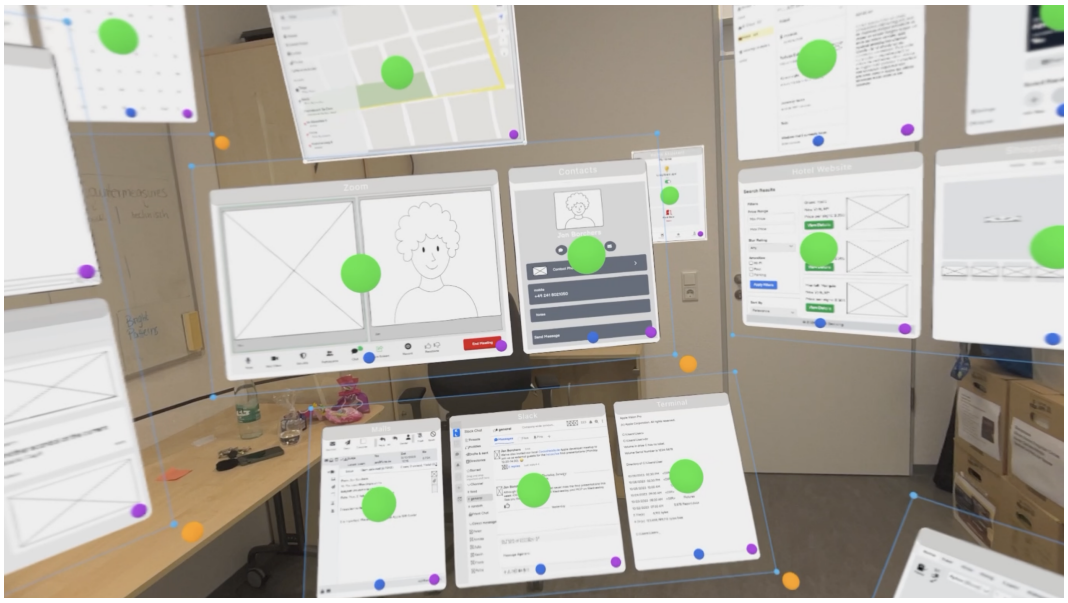


Figure 4.6: A participant's workspace with all 20 study windows during the search task in the *Grouped Placement* condition. The windows were the screenshot representations carried over from the low-fidelity study. Several user-created groups with bounding boxes were visible. The green button on each window appeared during the search-task phase. Participants finished a trial by tapping the button on the target window.

quickly. Since rapid full-field flashing is uncomfortable to watch and can be harmful to photosensitive users, we replaced the teleport with a short glide animation (0.18 s), which removed this effect. As the prototype changed after the pilot, the pilot session has been excluded from all analyses. The 12 participants of the evaluation study interacted only with the refined version.

Chapter 5

Prototype Evaluation

The formative study answered what users expect an AR window layout to do. Chapter 4 translated those expectations into a working prototype with three conditions. This chapter evaluates that prototype. Whereas the first study elicited mental models, this study tests whether the derived designs hold up in use. It also assesses whether the structure users asked for actually improved how they selected, found, and experienced windows in AR. The study materials, collected data, analysis scripts, and prototype source code are available in the supplementary material[1].

5.1 Methodology

Five hypotheses on tiled versus floating, tested via interaction logs and questionnaires.

Goals and Hypotheses The formative study answered RQ1 and RQ2 (Section 3.3). With the second study, we address RQ3 and RQ4 (Section 5.4). We wanted to identify whether the constrained conditions improved the usability of window management over free placement and which approach best matched users' preferences. Following prior work on virtual displays [Pavanatto et al., 2025a], we referred to the two group-based conditions (*Grouped Placement*, *Scrollbar*) as *tiled* and to the *Free Placement* baseline

[1] `https://hci.rwth-aachen.de/thesis/viewport-study/`

as *floating*. From RQ3 and the literature, we derived five hypotheses:

H1: Tiled canvases improve window-selection accuracy compared to floating windows.

H2a: Tiled canvases reduce multi-window search-task completion time compared to floating windows.

H2b: The scrollbar grouping reduces search time compared to static grouping.

H3: Tiled canvases decrease perceived cognitive load during multi-window workflows compared to floating windows.

H4: Tiled canvases are rated as more usable than floating windows.

H1, H2a, and H2b were measured from interaction logs produced by the prototype. H3 and H4 were measured through standardized questionnaires, NASA RawTLX and System Usability Scale (SUS). RQ4 was addressed through condition rankings, the per-condition ratings, and the qualitative feedback.

5.1.1 Study Design and Conditions

We used a within-subjects, fully counterbalanced design where the conditions differ minimally.

The study was a within-subjects experiment [Lazar et al., 2017] with a single factor, the window management condition, and three different conditions. The conditions were called *Free Placement*, *Grouped Placement*, and *Scrollbar*, as designed and implemented in Chapter 4. Every participant experienced all three conditions. As in the first study, we used complete counterbalancing over the 6 condition orders [Bradley, 1958; Williams, 1949]. We chose to have 12 participants, so that each order was used by two participants. The conditions differed minimally; each added exactly one mechanism over the previous one, so that differences between adjacent conditions were attributable to that mechanism (Section 4.1).

5.1.2 Apparatus and Setting

Lab session with the *Apple Vision Pro* and creating a baseline with a standardized instruction video.

The session took place in a room at our lab. The room measured 3.65m × 4.90m × 2.75 m (width × length × height). The room was kept identical across all sessions: The furniture was not moved, the blinds were closed, and the ceiling lighting was on, so that every participant encountered the same physical environment. Participants completed the questionnaires themselves on an *M1 MacBook Pro*, using a locally hosted web form that reproduced the original item wording and response scales. The view inside *Apple Vision Pro* was mirrored to an *M1 iPad Pro*, allowing us to follow the session and providing the screenshots used in this thesis. The recordings were not analyzed. Participants wore a first-generation *Apple Vision Pro* running the prototype described in Chapter 4 on *visionOS 26.5*. The space in front of the participant was kept clear of furniture. Participants were free to sit or stand; three of the 12 participants stood, and the remaining nine used a freely rotating chair so that all of them could turn toward the windows placed behind them. All interactions used *Apple Vision Pro* gaze-and-pinch input. To standardize the introduction across participants, we prepared a short instructional video that we showed at the start of each session. The video addressed putting on the headset, eye and hand calibration, the gaze-and-pinch gestures, and handling the prototype, including how to proceed when a known bug occurs. This way, every participant received identical instructions regardless of prior Vision Pro experience, before entering the practice phase of their first condition block. The video can be seen here[2].

5.1.3 Task and Scenario

The procedure is to let users practice, follow a scenario task, and do a search task.

Each condition comprised three consecutive parts:

Practice Phase: Participants spent around 5 minutes getting used to the condition using test windows.

[2] https://www.youtube.com/playlist?list=PLGqlSpIifaUE, *last accessed August 12, 2026*

Scenario Task: Participants followed a scenario opening 10 of the 20 available windows and, after that, opening the remaining 10 windows available in a random order.

Search Task: After all windows were placed, a search task was initiated, where participants had to find all 20 windows.

Practice task Before each condition, participants familiarized themselves with the mechanics of the current condition using five nondescript, colored test windows. This step was done so that the study windows were not previewed (Section 4.2).

Scenario extended to all 20 windows, with none closed so search stays constant.

Scenario task Participants then built up a workspace following the same *weekend trip* scenario as in the first study, adapted in two ways. First, the scenario was extended to all 20 windows, structured in five phases:

- **Planning the trip:** Maps, Hotel Website, Calendar
- **Comparing options:** Shopping, Mails
- **Coordinating:** Weather, Zoom
- **Filling the workspace:** YouTube, Slack, Banking
- **Final open phase:** The remaining 10 windows were added in randomized order and placed freely.

Second, unlike the first study, no window was closed. Every search trial ran over an identical, full 20-window workspace, keeping search difficulty constant across participants and conditions. The scenario again allowed the layout to grow incrementally, as it would during real work, rather than presenting all windows at once (Section 3.1.4).

Search task After the scenario, participants completed 20 search trials per condition. Each trial displayed an overlay naming a target window; dismissing the overlay started the search. Participants chose the window by tapping the green search button in the middle of the window that they believed to be the target (Figure 4.6). This task followed the established paradigm of window-switching evaluations on head-worn and virtual displays, in which participants locate and select a named target window among many open ones [Ens et al., 2014; Pavanatto et al., 2025b]. Unlike these studies, however, our participants searched a workspace they arranged themselves so that retrieval could draw on the spatial memory formed during the scenario task. Each of the 20 windows was the target exactly once per condition. Trials were shuffled with the constraint that no two consecutive targets came from the same category. The overlay-reading time and the search time (overlay dismissal to correct selection) were logged separately to ensure that reading the prompt did not contaminate the search measure.

Search task on self-arranged layouts leveraging spatial memory.

5.1.4 Participants and Procedure

We recruited 12 participants through the same channels as in the formative study (Section 3.1.5); Table 5.1 summarizes the demographics of the second study. Prior participation was not an exclusion criterion. Eleven of the 12 participants had taken part in the formative study and were, therefore, already familiar with the *weekend trip* scenario and the general study format. Since the study required wearing an HWD, we additionally recruited only people who needed no vision correction or could wear contact lenses, as *Apple Vision Pro* cannot be worn reliably over glasses. Ten of the 12 participants studied or worked in computing-related fields (computer science, IT, or media informatics), and the remaining two came from materials science and construction robotics backgrounds.

Twelve participants did the study, of whom 11 also took part in the formative study.

When questioned again about their multitasking habits, participants described the same pattern as in the formative study (Section 3.2.1). They reported having a primary task

Same primary-center, secondary-periphery pattern as in the first study.

on a main display, with secondary content such as communication, music, or reference material at the periphery, often distributed across monitors and devices. Given that 11 of the 12 participants also took part in the first study, this consistency was expected.

Participants	12
Age	M=26.5, SD=3.0 (range 22–32)
Gender	5 male (42%), 7 female (58%)
Immersive experience	12 some (100%), 8 with *Apple Vision Pro* experience (67%)
Number of monitors	1: 4 (33%), 2: 5 (42%), 3+: 2 (17%), varies: 1 (8%)
Formative study	11 also participated (92%)

Table 5.1: The demographics of our second study with the high-fidelity prototype.

A session contains consent, demographics, three counterbalanced condition blocks, and a final ranking.

Each session began with a welcome, the consent form, and a short demographics questionnaire, followed by the instructional video. Participants then completed the three condition blocks in their counterbalanced order. Each block consisted of the practice phase, the scenario task, and the search task described above, followed by a layout Likert rating, NASA RawTLX, and SUS for that condition (Section 5.1.5). After the final condition, participants ranked the three conditions, answered the free-text questions, and were debriefed. A session lasted around 60 minutes, including breaks and the practice phase. A condition block took roughly 15 minutes, with the first block typically lasting the longest and the third the shortest as participants became familiar with the procedure. As in the formative study, participants were compensated with drinks and snacks.

5.1.5 Measures and Analysis

All analyses were carried out in a *Python* script. We collected usage data from the prototype and self-report data from the questionnaires:

- **Interaction logs**: Every session produced a timestamped event log (Section 4.2). From these logs, we derived, per trial, the number of wrong selections and whether the target was found on the first attempt (H1), as well as the search time from overlay dismissal to correct selection (H2a, H2b). The logs also captured counts of snapping, grouping, compacting, and scrolling events per condition.

- **Questionnaires**: After each condition, participants rated the layout on eight items using a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree). The items were identical to those in the formative study (Section 3.1.6), namely suitability for writing and research work, for creative and design work, and for communication and email; how quickly the layout was understood; the predictability of where new windows would appear; the sense of control over window placement; how mentally demanding the layout felt; and its overall usefulness for daily work. Using identical items allowed the ratings, in particular the participants' sense of control, to be compared across the two studies. Participants further completed the NASA TLX [Hart and Staveland, 1988], with each of its six dimensions rated on a 21-point scale (0–20). We analyzed the NASA RawTLX score, with the unweighted mean of the six dimensions rescaled to 0–100 [Hart, 2006].[3] Finally, participants completed the SUS [Brooke, 1995], which was scored following Brooke's procedure to the 0–100 range. The questionnaires are presented in Appendix D. After completing all three conditions, participants ranked them by overall preference.

- **Free-text feedback**: For each condition, participants described what they liked and disliked about that condition. These responses are analyzed in Section 5.3.

The statistical procedure mirrored the first study (Section 3.1.6). Each measure was compared across the three

[3] The questionnaire recorded the unscaled 0–20 mean; the rescaling to 0–100 was applied during analysis.

conditions with a Friedman test. We reported χ^2, p, and Kendall's W as effect size. To control for Type-I error, we applied Bonferroni corrections per measure family, $\alpha = .05/8 = .00625$ for the eight layout items, and $\alpha = .05$ for NASA RawTLX, SUS, and the ranking, as each was a single test. Significant omnibus tests were followed by Wilcoxon signed-rank tests on the three pairwise comparisons and Bonferroni-corrected ($\alpha = .05/3 = .0167$) with the matched-pairs rank-biserial correlation r as effect size. Search times were aggregated in two steps. Within each session, a participant's 20 search trials were collapsed to their median. This step created one value per participant and condition ($12 \times 3 = 36$ values), on which the Friedman test operated with $N = 12$. For the tiled-versus-floating hypothesis (H2a), each participant's two tiled medians (*Grouped*, *Scrollbar*) were additionally averaged before a pairwise Wilcoxon test against *Free Placement*, so that the tiled conditions were not double-weighted.

Friedman with Bonferroni-corrected Wilcoxon post-hocs; search times collapsed to per-participant medians.

5.2 Results

In this section we discuss the findings of the second study.

5.2.1 Manipulation Check

The interaction logs confirm that the conditions differed in practice. *Free Placement* sessions contained no tiling events at all. *Grouped Placement* and *Scrollbar* sessions showed frequent snapping and grouping activity (on average, 15.6 and 14.8 snap events and 5.8 groups formed per session). The compact button was used 0.8 times per session on average. The scrolling mechanic, however, was used sparsely (Figure 5.2). Five of 12 participants never scrolled in the *Scrollbar* condition (median 1.5 scroll events per session, maximum 23). This finding limits the extent to which H2b can speak to scrolling as such because, for nearly half the sample, the *Scrollbar* condition was effectively just *Grouped Placement*. We return to this point in the discussion.

Manipulation check: Grouping behavior confirmed while Scrollbar largely ignored.

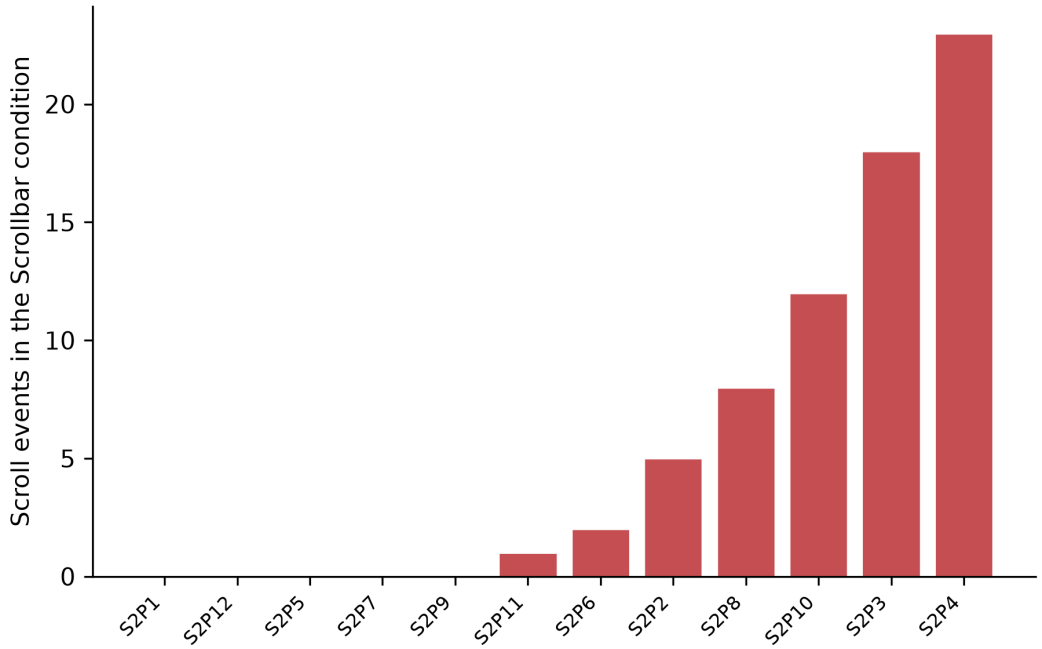


Figure 5.2: Manipulation check for the scrolling mechanic: Number of scroll events per participant in the *Scrollbar* condition. Five participants did not use the scrollbar while doing the task.

5.2.2 Selection Accuracy (H1)

H1 is not supported but ceiling shows study is insensitive.

Selection accuracy was at ceiling in all three conditions. Participants found the target on the first attempt in 99.4% of the trials, with only four wrong selections across all 720 search trials. With virtually error-free performance across all conditions, the data could not differentiate the conditions. This finding shows that H1 is not supported, but the ceiling means the study was insensitive to accuracy differences rather than evidence of their absence. We attribute the ceiling to the search buttons being large, unambiguous targets and to all 20 windows remaining visible or reachable by turning (see Section 5.4).

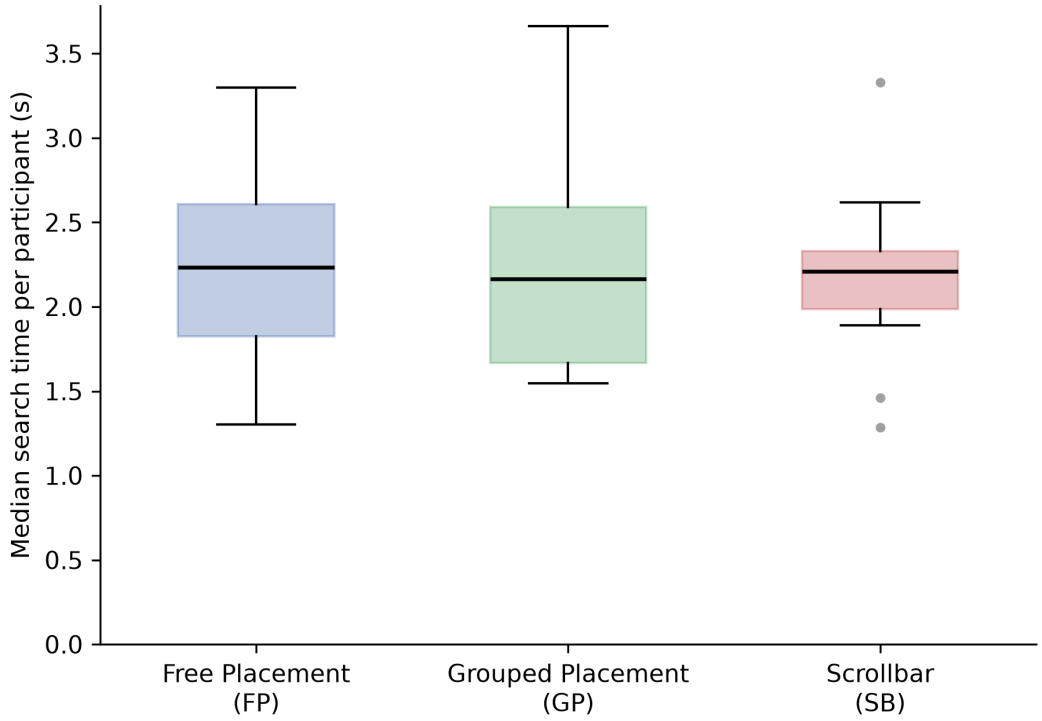


Figure 5.3: Median search time per participant and condition (FP = *Free Placement*, GP = *Grouped Placement*, SB = *Scrollbar*; $N = 12$). Boxes show the median and interquartile range over the per-participant medians. Grey dots mark outliers beyond them. No pairwise comparison was significant.

5.2.3 Search Time (H2a, H2b)

Search times nearly identical with no structure advantage.

Median search times were nearly identical across all conditions (Figure 5.3). 2.23 s in *Free Placement*, 2.17 s in *Grouped Placement*, and 2.21 s in *Scrollbar*. These values are medians. A Friedman test on the per-participant medians showed no effect of condition ($\chi^2(2) = 1.50$, $p = .472$, $W = .06$). The planned tiled-versus-floating contrast (H2a) was likewise non-significant (Wilcoxon $p = .622$, $r = .18$), as was the contrast between the two tiled conditions (H2b: $p = .266$). Neither H2a nor H2b is supported. This finding shows us that with 20 windows arranged by the participants themselves, structure provided no measurable search-time advantage over *Free Placement*.

5.2.4 Perceived Workload (H3)

Workload moderate and statistically indistinguishable across conditions.

NASA RawTLX scores did not differ significantly between conditions ($\chi^2(2) = 2.09$, $p = .353$, $W = .09$). Descriptively, *Grouped Placement* produced the lowest workload (M = 34.7, SD = 17.8, on the 0–100 scale), followed by *Scrollbar* (M = 37.8, SD = 18.0) and *Free Placement* (M = 40.4, SD = 14.3). Workload was thus moderate in all three conditions, and H3 is not supported.

5.2.5 Usability (H4)

Grouped rated highest, Scrollbar lowest and near the benchmark.

SUS was the one standardized measure on which the conditions differed significantly ($\chi^2(2) = 11.87$, $p = .003$, $W = .49$; Figure 5.4). *Grouped Placement* received the highest score (*Mdn* = 83.8, M = 81.9, SD = 13.3), above the commonly used benchmark of 68 [Bangor et al., 2008; Sauro and Lewis, 2016],[4] followed by *Free Placement* (*Mdn* = 82.5, M = 78.3, SD = 15.9) and *Scrollbar* (*Mdn* = 72.5, M = 68.3, SD = 20.0). Post-hoc tests located the difference within the tiled conditions. *Grouped Placement* was rated significantly more usable than *Scrollbar* ($p = .001$, $r = 1.0$),[5] with 11 of the 12 participants rating *Grouped Placement* higher and none rating it lower, while neither tiled condition differed significantly from the floating baseline (*Grouped* vs. *Free*: $p = .202$, $r = .46$, favoring *Grouped*; *Scrollbar* vs. *Free*: $p = .062$, $r = .62$, favoring *Free*). H4 is, therefore, not supported. Rather than tiling itself affecting usability, the effect depends on the specific tiling implementation. User-instantiated grouping achieved usability comparable to the floating baseline, whereas introducing the scrolling mechanism significantly reduced it. Notably, this penalty was not confined to participants who used the mechanism. Since only one participant rated the two conditions equally, exactly four of the five participants who never scrolled

[4] The benchmark of 68 is a descriptive reference point drawn from large SUS samples; with $N = 12$ and standard deviations of 13–20, our condition scores should not be read as reliably above or below it.

[5] The matched-pairs rank-biserial correlation reaches its maximum whenever all discordant pairs run in the same direction, independently of how large the individual differences are.

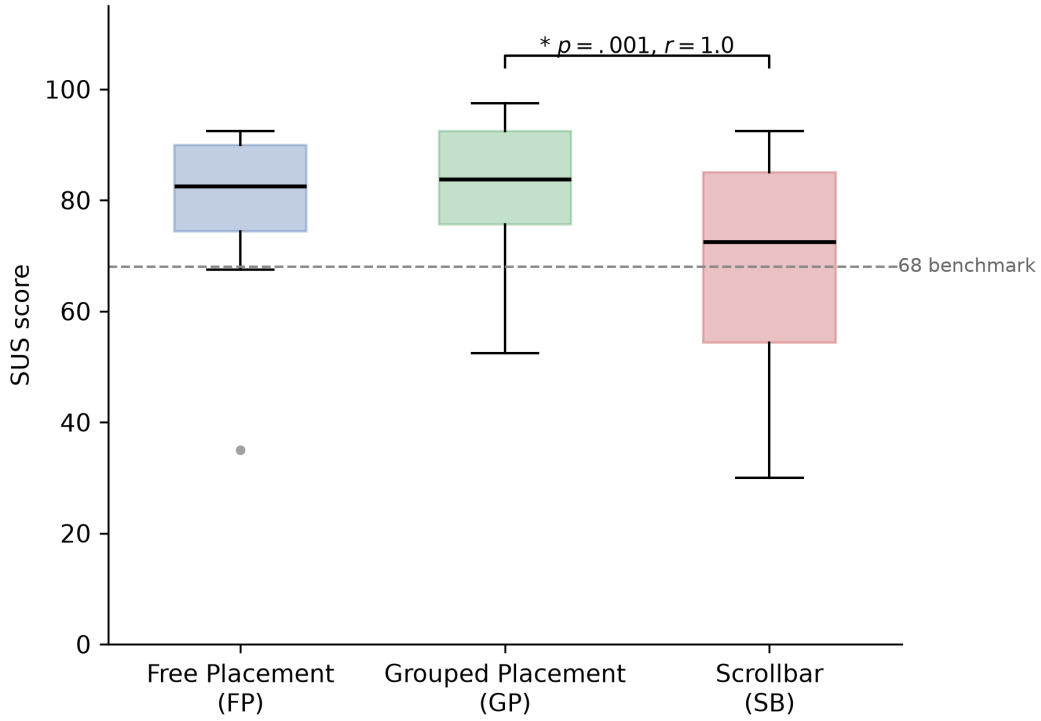


Figure 5.4: SUS scores per condition ($N = 12$). Boxes show the distribution across participants. The dashed line marks the commonly used benchmark of 68. The bracket indicates the only pairwise difference significant after Bonferroni correction (Grouped vs. Scrollbar, $p = .001$, $r = 1.0$). Grey dots show outliers.

(Figure 5.2) still rated *Scrollbar* lower than *Grouped Placement*. The reduction is therefore better attributed to the scrollbar as an affordance, introducing some combination of visual clutter, accidental activation, and an unexplained control, than to the scrolling interaction itself.

5.2.6 Layout Ratings

Only *understood* differed on rankings, control parity with opt-in structure.

Of the eight layout items, one differed significantly after Bonferroni correction, namely how quickly the layout was *understood* ($\chi^2(2) = 10.67$, $p = .005$, $W = .44$), favoring *Free Placement* (M = 4.75) over *Grouped Placement* (M = 4.42) and *Scrollbar* (M = 4.00). Post-hoc, only the *Free Placement*–*Scrollbar* contrast survived correction ($p = .008$, $r = 1.0$;

every participant whose ratings differed rated *Free Placement* higher). The remaining contrasts showed large rank-biserial effects but could not reach corrected significance because too few participants differed at all. For *Grouped* versus *Scrollbar*, only four participants gave different ratings (all favoring *Grouped*), for which the exact Wilcoxon test's smallest attainable p was .125. No other item reached significance; suitability for writing and research ($p = .061$, $W = .23$), suitability for communication ($p = .060$, $W = .24$), and daily work ($p = .089$, $W = .20$) came closest, all three descriptively favoring *Grouped Placement* (M = 4.50, 4.25, and 4.17, respectively). Notably, the *sense of control* did not differ between conditions (M = 3.92 / 4.25 / 4.00 for *Free Placement*, *Grouped*, and *Scrollbar*; $\chi^2(2) = 1.72$, $p = .423$, $W = .07$). Unlike the imposed constraint of the first study, which sharply reduced felt control (Section 3.2.2), the user-instantiated structure of the prototype carried no control penalty. We return to this contrast in the discussion.

5.2.7 Rankings and Order Effects

Grouped Placement ranked first descriptively with no order effects found.

Asked to rank the three conditions, seven of 12 participants ranked *Grouped Placement* first, three *Free Placement*, and two *Scrollbar*. The Friedman test on rank scores was not significant ($\chi^2(2) = 4.50$, $p = .105$, $W = .19$), so the preference for *Grouped Placement* is descriptive rather than statistical at this sample size. No order effects were observed. Analyzing the eight layout items, NASA RawTLX, and SUS by presentation order revealed no significant effects (all $p \geq .096$), and we also found no evidence of order effects.

5.3 Qualitative Results

Free-text answers were open-coded into 72 codes and organized into five themes.

After each condition, participants answered open questions on what they liked, disliked, and would change. After all conditions were completed, they justified their ranking, named the most natural condition, and described their ideal interaction. These free-text answers were analyzed following the same reflexive approach as in the first study

(Section 3.1.6). The answers were open-coded by the author using *Taguette*[6], yielding 66 codes over 192 coded segments, which we organized into five themes:

- Grouping as the preferred structure
- Freedom as the baseline expectation
- The rejection of the scrolling mechanism
- The demand for flexibility inside structure
- The persistence of desktop concepts

As the answers were short written statements rather than interview transcripts, the themes are closer to the data than in the first study. When the data continue a theme from the formative study, we reference it directly. Quotes are attributed as S2P1–S2P12. Quotes that are marked *(Translated)* were translated from German.

5.3.1 Grouping as the preferred structure

Ten of 12 praised grouping while 7/12 called it the most natural condition.

Grouping was by far the most frequently praised mechanism. Ten of 12 participants described it positively across all three conditions and seven named *Grouped Placement* the most natural condition. The reasons participants gave point directly at the semantic grouping model of the first study (T2). Participants experienced the mechanism as an externalization of a structure they already carried in their heads:

> *"The grouped placement condition felt most natural because I tend to build groups anyway (even in the free placement condition), so it was a natural extension."*
>
> —S2P8

[6] `https://www.taguette.org`

> *"I grouped windows into categories in my mind anyway so having that also represented in the app felt very intuitive"*
>
> —S2P10

> *"Some windows just belong together semantically or for a certain task [...] Grouping them then felt natural"*
>
> —S2P1

Grouping predates the prototype and was missed when unavailable.

That these categories predate the prototype is visible in the pre-study questionnaire. When asked about their desktop multitasking habits before the first condition, S2P2 already described an explicit zone system, with a central focus zone, an active zone for communication, and a passive zone for ambient content. This system is the same semantic organization participants later constructed with groups. The pull of grouping was strong enough to appear in its absence. After the *Free Placement* condition, five participants stated that what they would change in that condition was adding grouping:

> *"I would like the grouping feature back in some way, I did miss it from previous layouts"*
>
> —S2P10

> *"grouping would have been nice"*
>
> —S2P12

S2P11 gave a more specific reason for wanting to align related windows when it was not possible:

> *"Some windows just fit together perfectly, and I wanted to align them perfectly side by side (edge to edge), but it didn't work."*
>
> —S2P11 Translated

5.3.2 Freedom as the Baseline Expectation

The structured conditions carried no control penalty.

At the same time, 10 participants emphasized free control over window placement in this prototype:

> *"Freedom!!"*
>
> —S2P4

> *"There were no guardrails that kept me from placing the windows how I wanted them to be placed"*
>
> —S2P2

Crucially, these statements were not reserved for the *Free Placement* condition; participants made them about the structured conditions as well:

> *"I could just put windows wherever I wanted even in "strange" places. The grouping was very convenient and made finding things easier"*
>
> —S2P6, Condition *Scrollbar*

> *"I liked that I was able to place the windows freely wherever I want"*
>
> —S2P3, Condition *Grouped Placement*

S2P2 summarized this sentiment for all conditions:

> *"they all felt relatively the same in therms [sic] of natural feel since non of them really restricted any placement."*
>
> —S2P2

Opt-in structure did not register as a constraint; some wanted more of it

This finding is the qualitative counterpart to the control parity in the ratings (Section 5.2) and to the first study's

central distinction between imposed and self-chosen structure (T5). Since grouping was opt-in, structure did not register as a constraint. Participants even asked for *more* opt-in structure. For example, S2P5 wanted the snap indicator added to *Free Placement*, and S2P11 liked the placing indications:

> *"I thought it was great that I could see pretty clearly where my window was positioned thanks to those red bars."*
>
> —S2P11 Translated

A minority of four nevertheless ranked *Free Placement's* predictability above grouping's convenience, citing:

> *"The free placement felt most natural to me, because I felt like I was in control throughout and there were no unexpected behaviors."*
>
> —S2P3

S2P4 described this trade-off:

> *"Even though grouped placement is actually more convenient to use, the free placement was the most comfortable one for me, since it forced me to spread more throughout the room, which lead to a better use of the space to retrieve and control the windows."*
>
> —S2P4

Spatial spread as retrieval, not as missing structure.

S2P4's reasoning connects two threads from prior work, namely, the benefit of physically navigating spatially distributed content [Ball and North, 2008; Ball et al., 2007] and the role of stable window locations for spatial memory [Tak et al., 2009]. For this participant, spreading windows through the room was not a cost of missing structure but the retrieval mechanism itself.

5.3.3 The rejection of the scrolling mechanism

Four of 12 rejected scrolling, 4/12 ignored it. Criticism targeted the mechanism, not the idea.

The scrolling mechanism drew the clearest negative reactions from our participants. Four participants actively rejected it:

> *"I did not like the scrollbar at all. It felt unnecessary and only cluttered the screen. When trying it out, the scrolling itself felt abrupt and more confusing than just using the two windows next to each other."*
>
> —S2P1

> *"I don't need it. I don't want it."*
>
> —S2P7

Four other participants ignored it:

> *"[...] I ignored the scrolling completely [...]"*
>
> —S2P2

Scrolling rarely used, sometimes accidental, with an untested memory cost.

These findings matched the interaction logs, in which five participants never scrolled (Figure 5.2). Two participants triggered scrolling accidentally (S2P4, S2P8). Furthermore, S2P8 articulated a plausible mechanism by which scrolling could undermine spatial retrieval, which the search-time data did not detect, possibly because scrolling was too rarely used to disturb layouts:

> *"I did not really like the scrollbar, I did not intentionally use it and triggered it on accident a couple of times. I also felt like I was worse at remembering the placement of the windows, as I would usually know where I had placed a group but not which group a window belonged to."*
>
> —S2P8

S2P9 made explicit the mismatch with the underlying mental model:

> *"I am not able to scroll in the real world. I just look at where I put the things."*
>
> —S2P9

The implementation was rejected, not the concept.

Notably, most criticism targeted our implementation of the scrollbar rather than scrolling as such. Participants proposed changes, including vertical scrolling (S2P7, S2P9), restricting scrolling to one-dimensional groups (S2P7), and better usage indication (S2P11). Some participants also treated it as a harmless extra feature:

> *"Scrollbar and Grouped are basically the same to me but more features that don't hinder usage are always better."*
>
> —S2P12

S2P9 ranked Scrollbar first precisely because it was an opt-in feature:

> *"The only reason Scroll bar is first, is because it isn't mandatory. It's basically Free Placement with the option to group and scroll."*
>
> —S2P9

5.3.4 The demand for flexibility inside structure

Participants accepted the group but resisted rigidity within it.

Participants accepted groups as structure but pushed back against rigidity *within* them.

> *"If anything, I disliked how rigid the snapping was. I would prefer the windows inside a group not to be arranged in a grid but more flexible."*
>
> —S2P1

S2P8's ideal system was having options for rearrangement:

> *"I think the grouped placement comes closest, only missing the feature of rearranging windows inside the groups."*
>
> —S2P8

The slight automatic shift when adding a window violated expectations:

> *"When placing a window into a group, I would have expected for the currently existing windows to stay at their exact position."*
>
> —S2P3

Bounded freedom refined: Structure must stay flexible inside and not only be opt-in.

Finally, for S2P7, the shifting caused accidental group additions because the snap indicator went unnoticed while windows moved. Two participants (S2P3, S2P12) wanted each window inside a group to keep facing the viewer individually, as ungrouped windows did in the prototype, rather than all of them sharing the group's orientation. With several windows side by side, the outer ones then sat at an angle to the user instead of turning toward them, as they would on a curved monitor. Read in the context of the first study, this finding refines the notion of bounded freedom: Acceptable structure is not only opt-in at the level of the mechanism but must also preserve placement freedom inside the structure it creates.

5.3.5 The persistence of desktop concepts

Participants kept using desktop vocabulary to describe things.

As in the first study (T4), participants reached for desktop concepts when describing improvements and ideal systems. Five participants wanted freely resizable windows instead of the three fixed sizes, whereas four asked for minimizing or closing windows:

> *"Not being able to minimize windows at all. Some windows are one and done and just clutter my view. I am done shopping so I want to close it for example."*
>
> —S2P10

Moreover, three participants proposed tab or taskbar mechanisms to collapse groups:

> *"I would like the windows to be collapsable into a single tab or something and maybe even have a kind of taskbar that I can pop windows in and out of, but with grouping."*
>
> —S2P2

Desktop and mobile remained the vocabulary, even after three AR-native conditions.

S2P6 wished for a "multiple view mode" (S2P6), switching between task-specific arrangements. S2P11 had a similar idea where a button press would switch from work windows to music and video. S2P10 transferred a mobile pattern instead, proposing drag-based group folders:

> *"I am thinking like how you can group apps into folders on iPhone."*
>
> —S2P10

The participants, thus, thought in and used desktop and mobile window management vocabulary, even after experiencing three AR-native conditions.

5.3.6 Implementation Feedback

Spawn placement was the dominant usability complaint, separate from the conditions.

Finally, a set of recurring comments concerned the prototype rather than the window management conditions, and we report them separately in Section 5.4. The dominant complaint, which five participants raised, was about where new windows spawned. In this prototype, windows appeared at a fixed depth, sometimes behind existing windows in the focus zone:

> *"User may think [the] command did not work and may open it multiple times"*
>
> —S2P9

Prototype issues applied equally across conditions; reception was otherwise positive.

Individual participants further mentioned the small, unlabeled buttons, the fixed window sizes, and the inability to rotate windows away from the view. These issues applied equally to all conditions and therefore do not bias the comparisons, but they mark clear improvements for future iterations of the prototype. Despite these issues, the overall reception was positive. Three participants spontaneously described the study as fun (S2P6, S2P7, S2P10), "Good job on the implementation!" (S2P10). Moreover, the SUS scores of all three conditions lay at or above the conventional acceptability threshold [Bangor et al., 2008], although that of *Scrollbar* was only marginally so (68.3). This finding suggests that the null performance results reflect the conditions rather than a struggling implementation, at least for *Free Placement* and *Grouped Placement*.

5.4 Discussion

We discuss and interpret the second study of our *Apple Vision Pro* prototype.

In this section, we interpret the results of the evaluation study by drawing together the quantitative results (Section 5.2), the qualitative themes (Section 5.3), and the findings of the formative study (Chapter 3). We first revisit the hypotheses, and then we discuss the pattern that emerges behind them, answer RQ3 and RQ4, and derive implications for AR window management. We close with the limitations of this study. The contributions of the thesis as a whole and directions for future work are summarized in Chapter 6.

Hypotheses None of the directional hypotheses was supported as stated:

- H1: Selection accuracy was at ceiling.
- H2a/H2b: Search times did not differ between conditions.
- H3: Workload did not differ between conditions.

- H4: Tiling was not rated as more usable than the floating baseline.

Taken at face value, the constrained conditions did not outperform the baseline. However, when interpreted alongside the findings from the first study, a more nuanced pattern emerges.

Performance parity, structure valued for effort reduction and not speed.

Performance parity, preference divergence Structure did not affect objective performance; participants completed tasks just as quickly regardless of layout. Its effects instead emerged in subjective measures, including usability, near-significant task-suitability ratings, and the descriptive preference for *Grouped Placement*. This point mirrors the classic finding that constrained tiled layouts and free overlapping windows differ less in performance than in preference [Bly and Rosenberg, 1986]. Moreover, this finding also matches our theme T5, where participants valued structure not because it improved speed, but because it reduced arrangement effort while preserving a sense of control.

Opt-in structure preserves control and imposed structure reduces it.

The control penalty vanished In the first study, the imposed band constraint sharply reduced the sense of control. In this study, user-instantiated grouping showed control parity with *Free Placement*. Together with the finding that user-defined constraints preserve agency while saving effort [Niyazov et al., 2023], this point supports the central claim of this thesis: Structure per se does not cost users their sense of control but imposed structure does. *Free Placement*, present in both studies, was itself rated lower on control here than in the low-fidelity study (3.92 vs. 4.67), so the two studies are compared through their within-study patterns rather than their absolute values.

Scrollbar was rejected and elicitation better predicted structures than mechanisms.

Scrolling implementation The scrolling mechanism, which was derived directly from participants' own carousel metaphor in the first study (T3), was the clear loser. Nearly half the participants never used it, and it significantly reduced usability relative to plain grouping. Elicited ex-

pectations evidently predicted which *structures* users accept (grouping, T2) better than which *mechanisms* they will adopt.

Search was insensitive because no condition manipulated visibility.

Why search was insensitive With 20 windows arranged by the participants themselves, self-created spatial layouts already supported a roughly two-second-long retrieval in every condition, consistent with T3's turn-toward-remembered-locations mechanism and with spatial-memory accounts of window retrieval [Tak et al., 2009]. Prior work suggests why the conditions could not be differentiated. In handheld AR, Hueber et al. [2023] found that search-and-select times depended strongly on how much of the scene is visible at once: Their narrowest-field condition was the slowest and required the most device movement, while their widest-field condition eliminated the search phase almost entirely once participants found a posture from which the whole scene fit on screen. Our conditions held this variable constant. All 20 windows remained open and reachable by turning in every condition, and grouping changed how windows were arranged rather than how many were visible. Regarding the dimension that governs AR search performance, our conditions were, therefore, identical, which is a more plausible account of a null finding than an absence of any effect of structure. Differences may emerge where structure does change visibility, for instance, with more windows than can fit in the space around the user. The ceiling in accuracy points in the same direction.

Grouped matched the mental models and user-instantiated structure welcomed.

Answering RQ3 and RQ4 RQ3 asked whether the constrained conditions improved the usability of window management over *Free Placement*. The answer was differentiated. Objectively, no: Selection accuracy, search time, and workload showed no advantage of structure. Subjectively, user-instantiated grouping matched *Free Placement* on usability, workload, and sense of control while being descriptively preferred. Furthermore, the near-significant suitability ratings pointed in its favor, whereas adding the scrolling mechanism significantly reduced usability. This finding showed that constraints help in the right circumstances;

structure that users instantiated themselves costs nothing and was welcomed. A mechanism that moved windows on the users' behalf was rejected. RQ4 asked which approach best matched users' mental models and preferences. Here, the evidence converged on *Grouped Placement*. Seven of the 12 participants ranked it first, and seven also identified it as the most natural condition. Qualitative accounts reinforced this finding by suggesting that *Grouped Placement* reflects a grouping structure that participants already maintain mentally (Section 5.3, T2). *Free Placement* remained the preferred option for the minority of participants who prioritize predictability and spatial spread over arrangement convenience, consistent with the freedom-versus-structure theme (T5) identified in the first study.

Four implications: opt-in structure, within-group flexibility, avoid auto-relocation, desktop familiarity.

Implications for AR window management Four implications follow for the design of window management on HWDs. First, structure should be strictly opt-in. Self-chosen structure carried no control penalty in either study, while the first study showed that imposed structure did, converging with work on user-defined constraints [Niyazov et al., 2023]. Second, structure must preserve freedom *within* itself. Participants accepted the group as a unit but rejected the rigid grid and the automatic re-layout inside it (Section 5.3). Third, mechanisms that relocate windows independently of the users' own actions should be treated with caution. Retrieval in AR rests on spatial memory of self-chosen locations [Tak et al., 2009]; scrolling broke precisely this mechanism:

> *"I also felt like I was worse at remembering the placement of the windows as I would usually know where I had placed a group but not which group a window belonged to."*
>
> —S2P8

Build on desktop conventions, not around them.

Our search-time data did not detect such a cost, but the mechanism was too rarely used to put it to a real test. Fourth, desktop concepts remained the vocabulary in which users thought (T4). Selectively recreating familiar mechanisms such as minimizing, resizing, and tab-like

collapsing is likely to ease adoption more than inventing AR-native replacements for them as of now.

Sample, task design, scrollbar exposure, and prototype specifics bound the conclusions.

Limitations Several limitations should be considered when interpreting these findings. First, the sample size was small. Although 12 participants allowed complete counterbalancing, it limited statistical power, as reflected by the exact-test floors in the layout ratings (Section 5.2). Consequently, null findings, particularly for workload and the ranking task, should be interpreted with caution. Second, the participant sample was young, predominantly drawn from computing-related fields and uniformly experienced with immersive technology, including eight participants with prior *Apple Vision Pro* experience. As a result, the findings are more representative of early adopters of technology than of the general population. In addition, 11 of the 12 participants had previously taken part in the formative study (Section 5.1.4). While this familiarity reduced learning overhead during the evaluation, it also meant that most participants had already been exposed to the formative study's conditions and concept videos, which may have influenced their expectations of the evaluated designs. Finally, the evaluation consisted of a single laboratory session per participant. As such, the reported results primarily reflect first-use impressions shaped by novelty and a relatively brief practice phase, and they do not support conclusions about long-term adoption, sustained use, or retention over multiple days.

Ceiling from fresh spatial memory, simple targets, and optimal arrangement.

The task design favored the ceiling effects we observed. Search trials began minutes after participants had arranged the 20 windows themselves; therefore, spatial memory was at its freshest. All windows remained open and reachable by turning, and the search buttons were large, unambiguous targets. Participants were moreover aware from the second condition onward that a search task would follow, and they arranged windows accordingly. This observation was also discerned by a participant, who stated that the search challenge "encouraged optimal placement" (S2P10). Thus, the search task was less a test of structure than a best-case test of spatial memory. Differences may well emerge with more windows, longer delays between arrangement

and retrieval, workspaces arranged by others, or targets requiring content inspection rather than recognition. The scenario windows themselves were static screenshot representations rather than live applications, so the workflows lacked content interaction, time-pressured window switching, and the dynamic rearrangement of windows characteristic of real work. The window set itself was informally sourced in the formative study rather than formally derived from a systematic task analysis. However, we note that no participant in either study questioned its selection, and one spontaneously remarked that the windows "represented my daily work quite well" (S2P10).

The Scrollbar was underused and its usability was tied to implementation, not the concept itself.

The *Scrollbar* condition was only partially used. Five of 12 participants never scrolled; therefore, for them, the condition was reduced to *Grouped Placement*, and H2b compared scrolling against static grouping. Its usability penalty was correspondingly an evaluation of this concrete implementation with its accidental activations and weak affordance rather than that of scrollable groups as a concept. Participants proposed improvements to the *Scrollbar* mechanism (Section 5.3). More generally, each condition was one implementation of its concept, and the SUS result itself demonstrated that implementation choices, not paradigms alone, drive usability judgments.

Finally, the qualitative data stemmed from short written answers rather than interviews and were coded by a single coder, creating themes closer to the surface of participants' reasoning than in the first study. The prototype ran on a single device with gaze-and-pinch as the only input. Prototype-level issues, such as the spawn placement of new windows (Section 5.3), while equal across conditions, affected the overall experience. These constraints do not undermine the main findings, but they limit the extent to which they can be generalized.

Chapter 6

Summary and Future Work

This chapter concludes the thesis with a summary of our contributions and a discussion of directions for future work.

6.1 Summary and Contributions

First, we elicit the mental model; then, we evaluate designs built from it.

This thesis asked how window and viewport management on HWDs should be structured once the physical screen is removed, along with the constraints that shaped four decades of window management. Rather than proposing a design and testing it, we inverted the order. We first elicited the mental models users bring to AR window management and only then built and evaluated a system derived from them. This two-phase approach was motivated by the observation that prior work offers many design solutions for placing content in immersive environments, but little evidence of what users actually expect such systems to do (Chapter 2).

Study 1: Users expect free placement with opt-in structure.

In the formative study (Chapter 3), 24 participants experienced three window management concepts, namely *Free Placement*, *Fixed Horizontal*, and *Elastic Horizontal*, on a low-

fidelity acrylic prototype, followed by interviews and a co-design phase. Regarding **RQ1**, the elicited mental model was spatial and semantic. Here, participants sized and placed windows by importance, grouped them by meaning with gaps as boundaries, retrieved them primarily by remembering where they put them, and transferred familiar desktop concepts into the new space. Concerning **RQ2**, participants wanted structure strictly on their own terms. Imposed constraints reduced their sense of control, whereas opt-in structure did not. We condensed this finding into the notion of *bounded freedom*, which is opt-in structure within free space.

A prototype turned these findings into three testable conditions on *Apple Vision Pro*.

We then translated these findings into the *SpatialWindowPlacer* prototype, a high-fidelity prototype for *Apple Vision Pro* (Chapter 4). Following the formative study's clearest signal, the prototype abandoned the globally constrained horizontal band in favor of a free, unbounded workspace in every condition. We reintroduced structure only as something users create themselves. The conditions progressively introduced one new mechanism at a time. First, free placement was introduced, followed by user-instantiated grouping with snapping, and, finally, an in-group scrollbar implementing the carousel-like scrolling participants had envisioned (T3).

Study 2: Performance parity, no control penalty for opt-in structure, and a rejected mechanism.

The evaluation study (Chapter 5) with 12 participants found no performance differences. Search times did not differ across conditions, and selection accuracy was at ceiling in all three conditions. Preferences, however, diverged. Regarding **RQ3**, user-instantiated grouping matched *Free Placement* on usability, workload, and sense of control, unlike the imposed constraints in the first study. By contrast, the scrolling mechanism significantly reduced usability and was ignored by nearly half the participants. Concerning **RQ4**, *Grouped Placement* best matched participants' mental models, with a majority ranking it first, though this preference was not statistically significant at this sample size.

Four contributions.

In summary, this thesis makes the following contributions:

- **An empirical account of users' mental models of AR window management:** This account is derived from

a low-fidelity elicitation and co-design study and condensed into five themes (Chapter 3).

- **The design and implementation of *SpatialWindowPlacer*:** This is a native visionOS prototype realizing three window management conditions, designed such that each introduces exactly one additional interaction mechanism (Chapter 4).
- **An empirical evaluation:** The evaluation shows performance parity across conditions, no control penalty for user-instantiated structure, and a descriptive preference for opt-in grouping over both free placement and a scrolling mechanism, which participants largely declined to use (Chapter 5).
- **Two methodological observations for elicitation-driven design**: Low-fidelity elicitation predicted which *structures* users would accept, but not which *mechanisms* they would adopt (Section 5.4). The study also contributes a reusable low-fidelity technique for comparing placement constraints, in which each condition's rule is enforced physically by the card backing rather than by instruction (Section 3.1.3).

6.2 Limitations and Future Work

The detailed limitations live with their studies; here we summarize what bounds the thesis as a whole.

The detailed limitations of the individual studies are discussed in their respective chapters (Sections 3.4 and 5.4). Here, we summarize what limits the thesis as a whole. Both studies drew on small, young, computing-heavy, XR-experienced samples, with eleven participants shared across the studies. Consequently, findings speak for early adopters in a single cultural context rather than the general population. The overlap carries a second cost beyond sample size: The same individuals whose mental models we elicited also evaluated the designs built from those models, so the evaluation may in part confirm what it was derived from. That the scrollbar was rejected despite originating from those participants indicates the studies were not merely self-confirming. However, the conclusion that the grouping matched participants' mental models would be

strengthened by replication with an independent sample. Both studies were single-session lab studies. The mental models we elicited are first-contact models, and the evaluation captures first-contact judgments. Long-term adoption or the evolution of strategies over weeks of real use cannot be inferred from either study. Each condition existed in exactly one implementation, and the evaluation itself demonstrated that implementation choices, and not paradigms alone, drive usability judgments.

Imposed versus chosen structure was never compared within one study.

A further limitation concerns the thesis's central comparison itself. Imposed and self-chosen structure were tested in separate studies; therefore, the contrast between them is confounded with prototype fidelity, task, and apparatus. The shared *Free Placement* baseline was itself rated differently across the two studies. The conclusion therefore rests on the pattern each study shows internally, supported by convergent evidence from work on user-defined constraints [Niyazov et al., 2023] rather than on a direct within-study manipulation of who defines the constraint. A study varying authorship while holding geometry fixed—for instance comparing a system-imposed band against a user-toggleable one—would test the claim directly.

Test with more windows, longer delays, and longitudinal real-world use.

These limitations point directly to directions for future work. First, for *scale and time*, the observed performance parity should be tested under more demanding conditions, including larger window counts, retrieval delays of hours or days, and workspaces arranged by others. These are the settings in which organizational structure, rather than spatial memory, would be expected to support retrieval. Longitudinal deployments with live applications rather than static windows would also reveal whether the preference for grouping persists in real work, despite the continual opening, closing, and rearranging of windows.

Redesign Scrollbar where we explore globe scrolling and global band at larger scales.

Second, regarding *mechanism redesign*, participants rejected this particular scrollbar rather than scrolling itself. They also proposed several concrete redesigns, including vertical scrolling, restricting scrolling to one-dimensional groups, and clearer affordances, which warrant evaluation before the carousel concept is discarded. Two design ideas that were deliberately deferred in Chapter 4 also belong here.

The first is whether scrolling the entire workspace around the user like a *globe* of windows can be made viable without causing the discomfort some participants anticipated. The second is whether the globally constrained band, abandoned after the formative study, might prove useful at scales where free arrangement breaks down.

Flexible within-group layouts and add minimize, tabs, continuous resizing.

Third, *richer structure* warrants research. The finding that structure must preserve freedom within itself invites designs with flexible within-group layouts, alongside the desktop-inherited features participants asked for, such as minimizing, tab-like collapsing of groups, free continuous resizing, and task-specific arrangements akin to virtual desktops in AR.

System-suggested groupings, explicitly confirmed to stay opt-in.

Fourth, regarding *automation under user control*, bounded freedom does not preclude system assistance. Instead, the system could, for example, suggest groupings based on semantic similarity, which users explicitly confirm before they are applied. Such an approach could reduce arrangement effort while keeping organizational structure opt-in, connecting naturally to work on user-instantiated constraints for layout optimization [Niyazov et al., 2023].

Test with diverse users, other headsets, input methods, and collaboration.

Finally, *broader populations and settings* warrant consideration. Future work should evaluate bounded freedom with participants beyond young computing students, including users who require vision correction as hardware support improves. It should also examine other HWDs, alternative input methods, and collaborative multi-user workspaces to determine how broadly the approach generalizes.

The screen's constraints should be an option, not a mandate.

Taken together, the two studies suggest a simple stance for the coming generation of AR window managers: Constraints inherited from the physical screen should be offered but not imposed. Users do not want the unbounded workspace structured for them; they want to decide when and how to structure it themselves.

Appendix A

Questionnaires Study 1

In the following, we include the questionnaire used in the formative study (Chapter 3). It was completed on a locally hosted web form in three parts. Part 1 was filled out before the first condition; a stop warning then prevented participants from continuing until all three conditions and their interviews were complete, after which they completed Parts 2 and 3. The rating and ranking sections therefore use the internal condition names *Infinite Scroll* and *Elastic Scroll*, which were withheld from participants during the conditions themselves (Section 3.1.2).

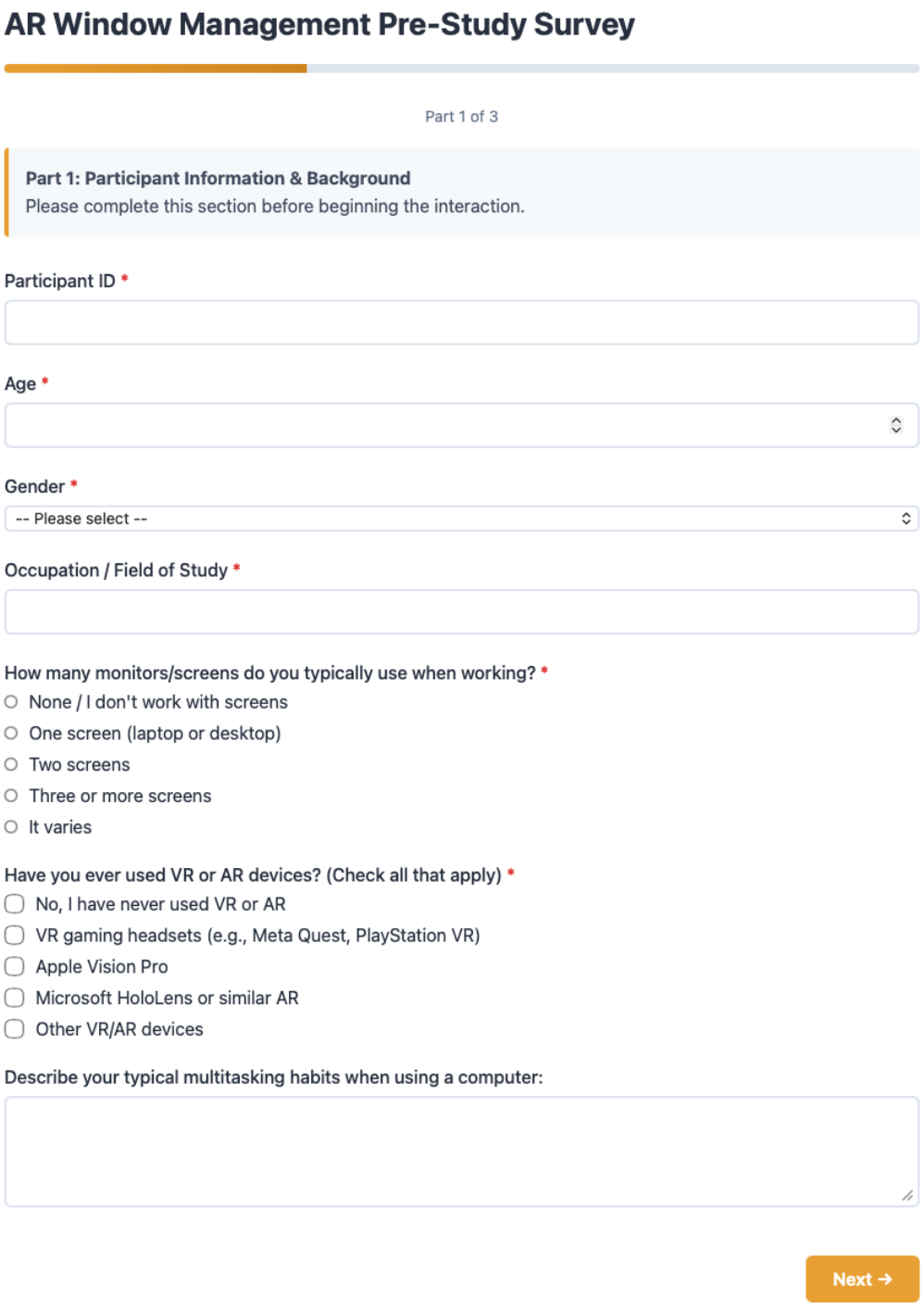


Figure A.1: Formative study questionnaire. Part 1 collected demographics, monitor setup, prior VR/AR experience, and a free-text description of everyday multitasking habits before the first condition.

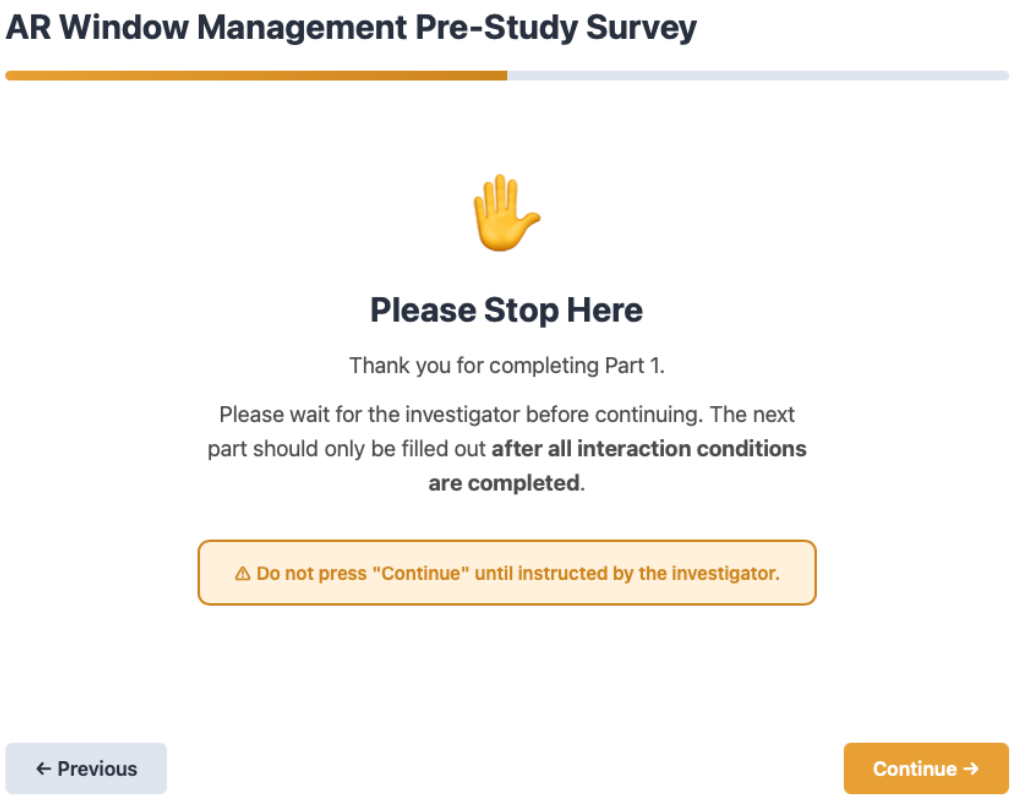


Figure A.2: A stop warning between Parts 1 and 2 prevented participants from proceeding until all three conditions and their interviews were complete, so the internal condition names used in Part 2 could not prime the elicitation.

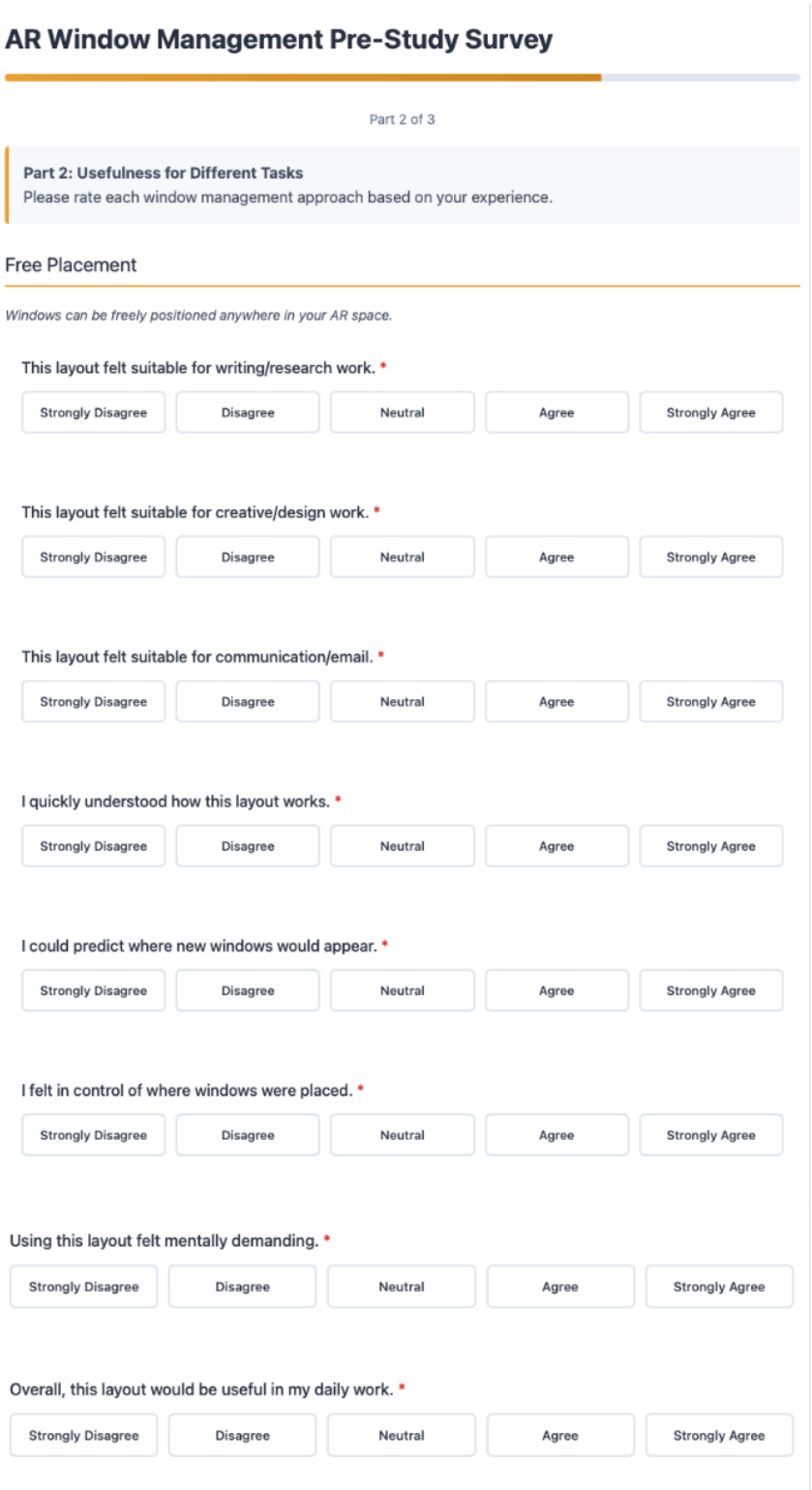


Figure A.3: Part 2 rated each condition on the eight items, shown here for *Free Placement*. The same eight items were used for all three conditions and for both studies.

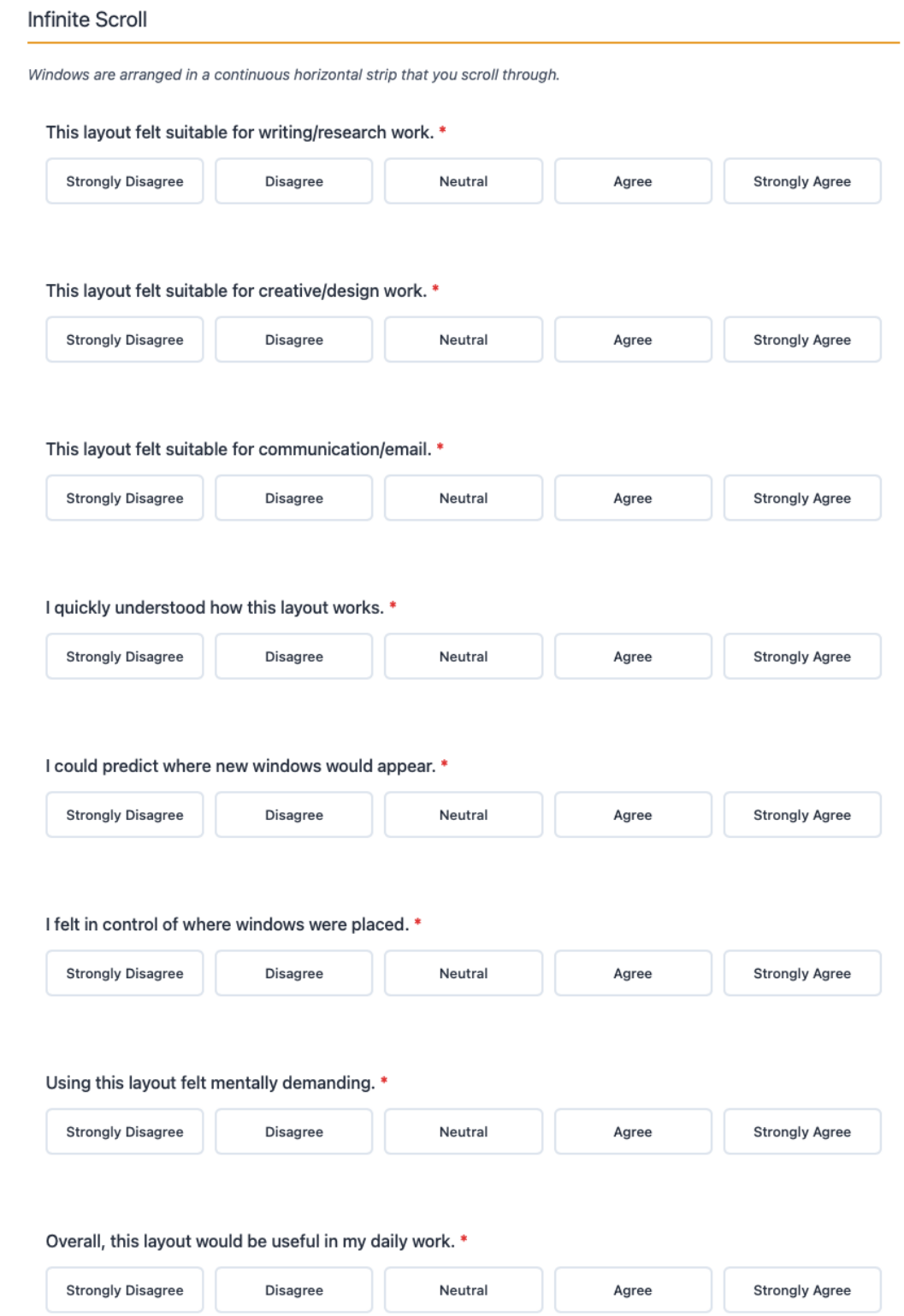


Figure A.4: The eight rating items for the *Fixed Horizontal* condition.

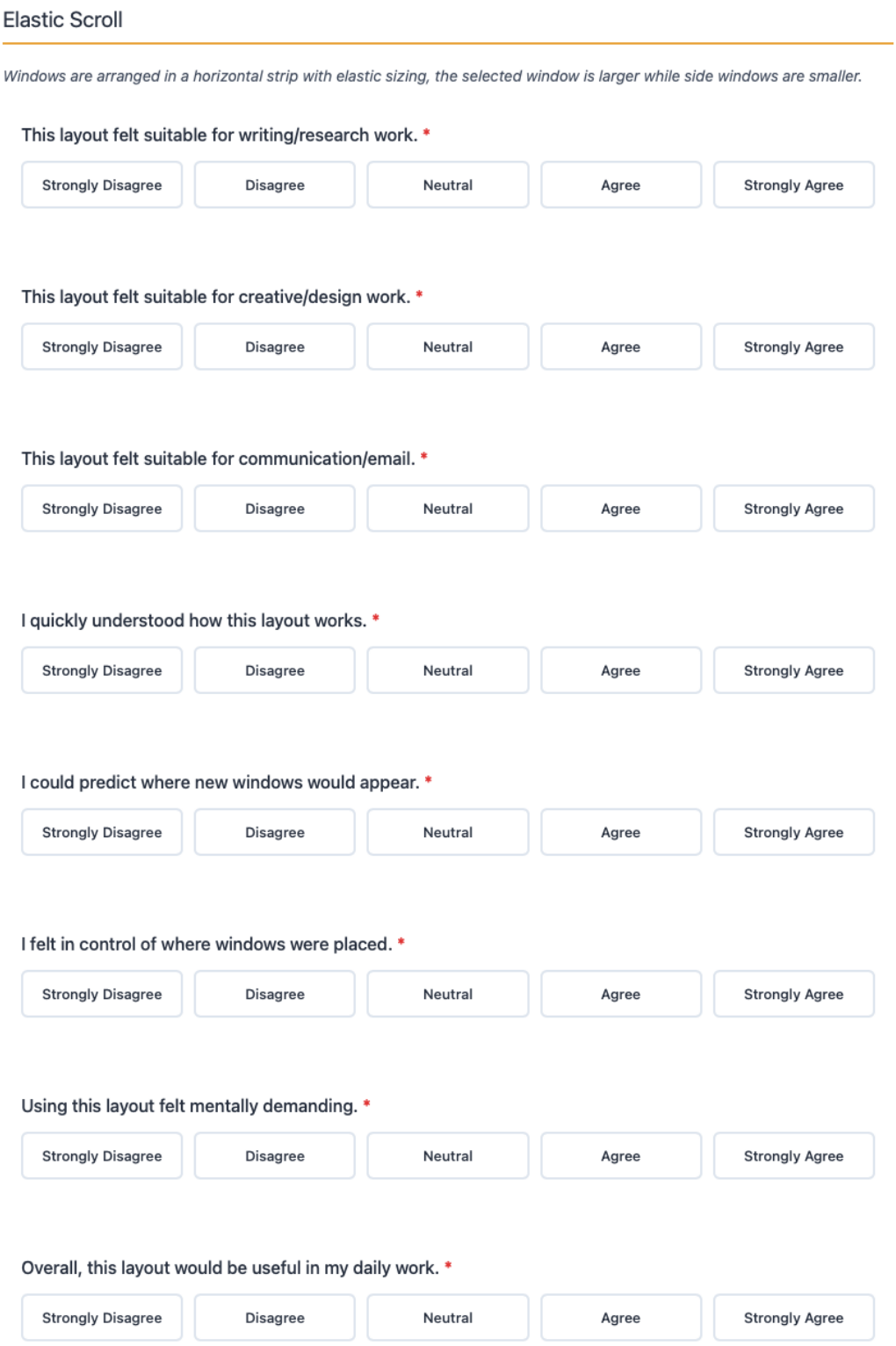


Figure A.5: The eight rating items for the *Elastic Horizontal* condition.



Figure A.6: Part 3 asked participants to rank the three conditions by overall preference.

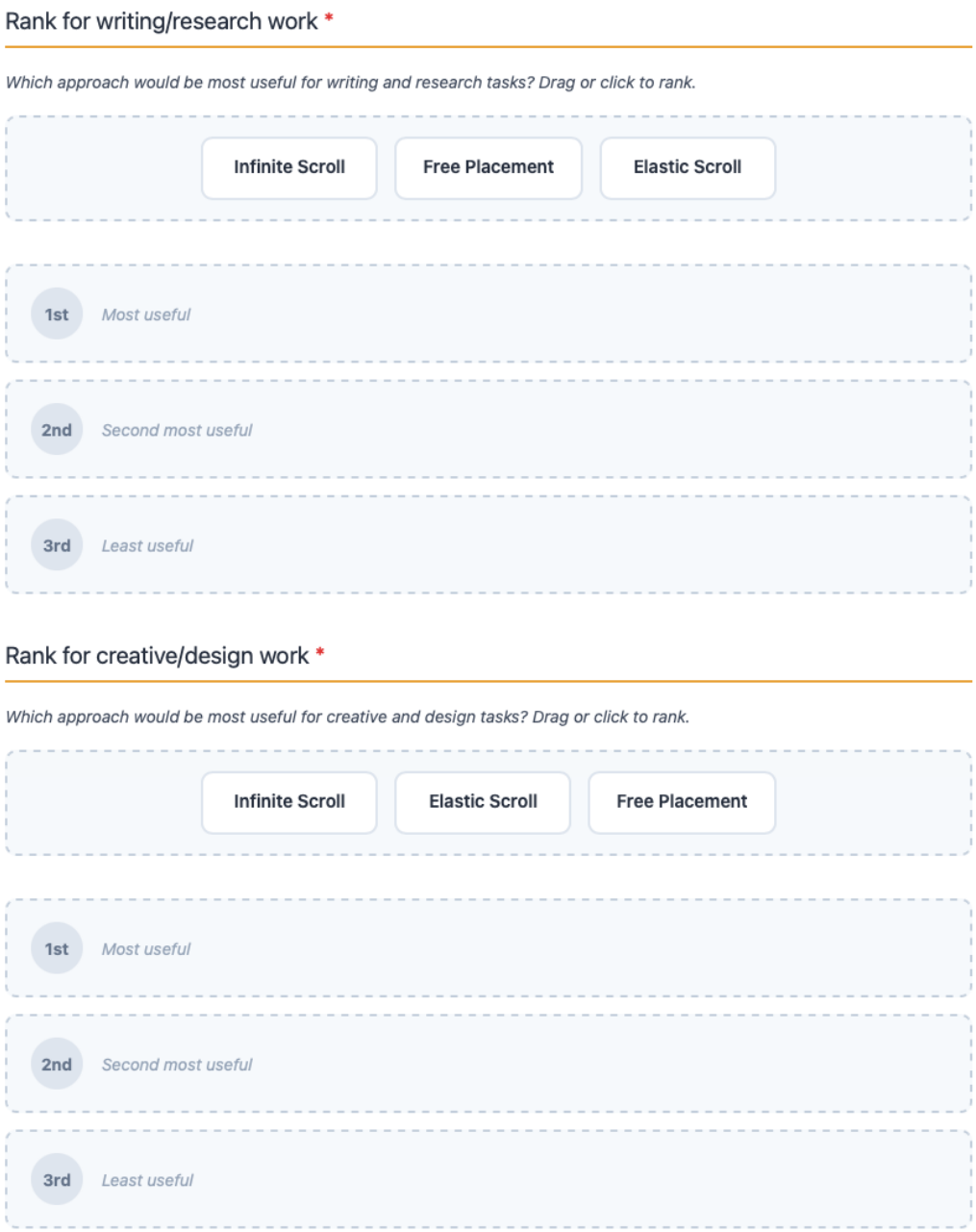


Figure A.7: Participants ranked the conditions separately for writing and research work and for creative and design work.

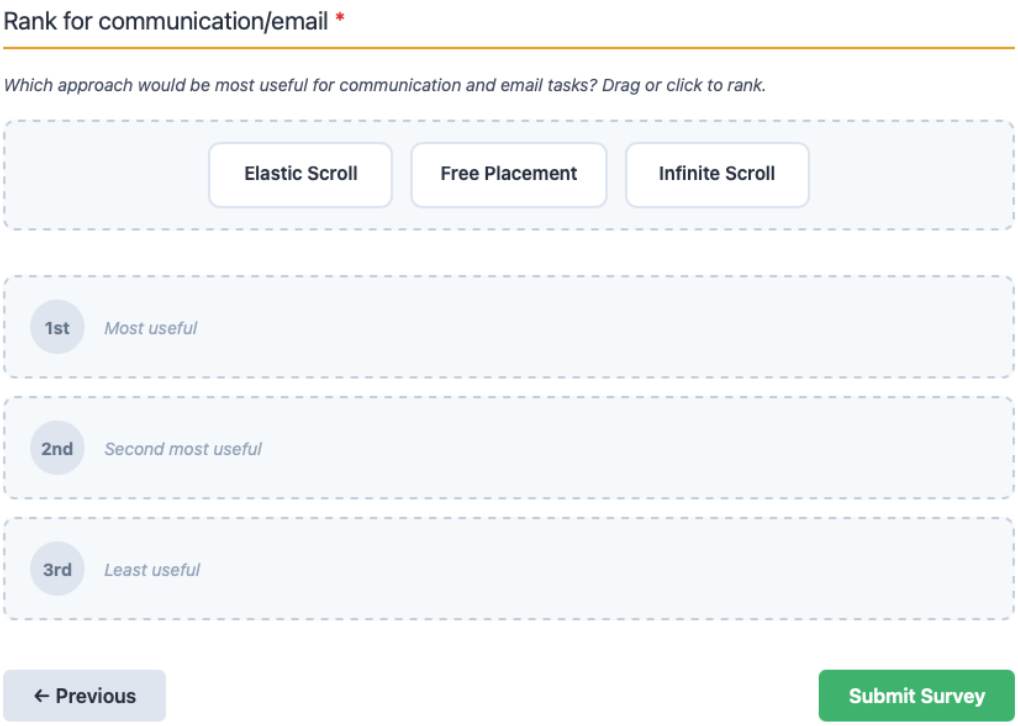


Figure A.8: The final ranking category, communication and email, completed the survey.

Appendix B

Codebook Study 1

The following codebook contains all codes constructed during the reflexive thematic analysis of the formative study, grouped under the five themes reported in Chapter 3.

Segment counts are reported for transparency only; following Braun et al. [2019], themes were judged by analytic meaning rather than by frequency. A segment may carry more than one code. The five themes are described in Table 3.7. Each theme is introduced by a row with an orange background. The final three groups fall outside the five themes: prototype and apparatus remarks, reactions to the concept videos, and the marker showing whether a segment originated in a drawing or an interview.

T1 Importance influences size, place and visibility

T2 Window organization

T3 Bringing windows into view

T4 Using desktop paradigms

T5 Freedom versus structure

Code	Segments	Description
T1 Importance influences size, place and visibility		
Size expresses importance	140	Sizing a window according to how much it currently matters. Includes explicit statements that the main task is largest.
Unneeded windows moved aside	71	Relocating a window out of the working area while keeping it retrievable. Distinct from minimising to a bar and from closing.
Main focus placed centrally	70	Positioning the currently most important window directly in front of the user. Covers where the focal window sits, not which window holds focus.
Unneeded windows minimized or iconized	45	Collapsing a window into a bar, icon, or widget rather than moving it in space.
Position expresses importance	40	Placing windows nearer to or further from the centre to signal priority, without reference to size.
Position is remembered and reused	37	Returning windows to the same spot across tasks or sessions, or expecting the system to restore a window's last position.
Focus designated by looking	32	Looking at a screen to work on it. This entails not moving it to a center.
Unneeded windows closed	31	Removing a window from the workspace entirely when it is no longer required.
Size follows content or context	20	Sizing driven by what the window contains, the device it imitates, or the space available, rather than by its priority to the participant.
Main focus switches depending on use case	20	There is not just one main window, but the focus can switch depending on factors.
T2 Window organization		
Grouping by meaning	215	Placing windows together because their content or task relates. Excludes proximity participants described as incidental.
Gaps as deliberate boundaries	108	Leaving space between windows to mark where one cluster ends and another begins. The gap is the structuring device, not leftover space. Excludes gaps participants described as unintentional.
Overlap used	80	Letting windows partially cover one another, including stacking and layering. Excludes transparency used to reveal what is behind.
Zones and allocated spots	60	Assigning a region of the space a standing role, independent of which windows currently occupy it. Excludes groups defined by their members.
Overlap rejected	32	Declining to overlap windows, or describing overlap as unacceptable under a given constraint.
Alignment and snapping	24	Lining windows up to shared edges, or expecting the system to align them on placement.
Groups manipulated as units	22	Groups that can be manipulated as a unit, by moving them around, minimizing them or changing the size all at once.
Enclosures and separators	13	Marking a group with a drawn or imagined border, box, or dividing line rather than with a gap.
Gaps rejected	8	Preferring windows packed together, or describing gaps as wasted space.
Gaps as incidental	7	Space between windows described as unintentional or as an artifact of arranging. Counterpart to Gaps as deliberate boundaries.
Switching tasks	3	Expression of how tasks are switched when focus is on another window.
Transparency to reveal what is behind	3	Participants wanting windows to be transparent to avoid hiding what is behind them.
Chronological ordering	2	Windows are ordered chronological in a FIFO principle.
Gaps persist when windows move	2	When moving around all windows gaps should persist and stay.
Limit on stacking depth	1	When stacking windows should not be stacked too deeply.
T3 Bringing windows into view		
Window opens in view	84	Expecting a newly opened window to appear where the user is already looking, at a predictable size. Excludes expectations that it is placed into free space instead.
Turning head or body to search	79	Locating an out of view window by physically turning rather than by scrolling or searching. Excludes remarks that turning is taking effort.
System moves windows automatically	61	Expecting the system to reposition existing windows on its own. Covers the expectation itself! Approval or rejection of it is coded separately.
Search and highlighting	57	Retrieving a window by naming or querying it rather than by moving through the space, including expectations that the match is highlighted.
Scrolling moves the whole workspace	56	Understanding the scroll as shifting all content at once rather than a single window or group. Input gesture and direction are coded separately.
Gaps close automatically	48	When closing a window a gap closes automatically, or when moving windows around.
Moving a window to the focus	44	Bringing a target to the user rather than the user to the target, by dragging, a button, or a gesture.
Scrolling as retrieval	41	Using the scroll to bring a specific known target into view, as opposed to browsing.
Closing window creates gap	35	Expecting the space a closed window occupied to remain visible as a gap. counterpart of Gaps close automatically.
Peripheral vision used	34	Keeping secondary content within side vision so it can be monitored without turning.
Window opens in free space	33	Expecting a new window to be placed into available space rather than at the focus, so as not to disturb the current task.
Scrolling follows the hand	32	The direction the scrollbar is moved with the hand, also moves the windows following the hand direction.
Scroll input expectation	24	Usage of scroll input, either with the scrollbar or with gestures.
Automatic movement rejected	23	Objecting to movement the participant did not initiate, whether inside a group or across the layout.
Automatic movement welcomed	21	Approving of system initiated movement, typically where it tidies the periphery rather than the focal window.

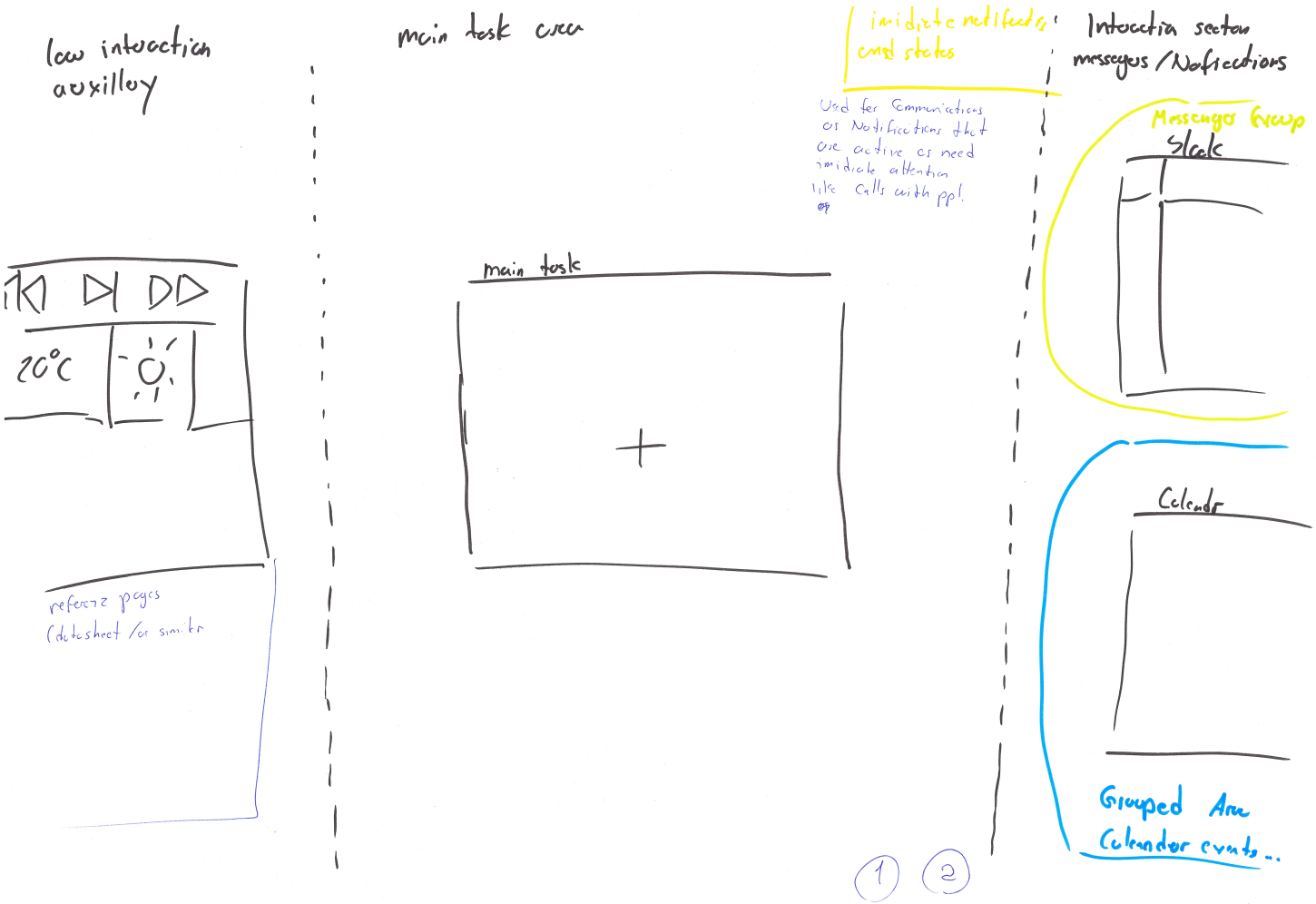
Code	Segments	Description
Scrolling rejected or doubted	19	Scrolling rejected by participants because of cybersickness, mental load, problems understanding how it works.
Repositioning windows by hand	17	Moving the windows around. Either by grabbing or other interaction methods.
Turning described as costly	11	Reluctance to look behind or turn repeatedly. The counterpart to Turning head or body to search.
Windows open at a default size	9	Windows having a default size that they open in.
Scrolling scoped to one window or group	9	Scrolling scoped to one window or a group of windows.
Scrolling moves opposite the hand	6	When scrolling the windows move opposite direction of hands.
Same application opened several times	3	The same window was opened several times to represent different instances of the same program.
Window opens in the main focus	2	Opening a window is opening the window in the main focus in front of view.
Gap closing on request	2	Gaps are closed not automatically but on request.
New window may interfere with the focus	1	When having several windows open they can interfere with main view.
Scrollbar size reflects window count	1	The scrollbar size is dependent on how many windows are open.
Two-handed scrolling	1	Both hands are used to move windows around.
T4 Using desktop paradigms		
AR mapped onto desktop setup	80	Describing the AR workspace by direct analogy to the participant's own monitor arrangement or desktop mechanisms.
Taskbar or dock expected	28	Wanting a persistent strip for opening, switching, or storing windows.
Menus expected	12	Wanting a menu or file-system structure for locating and opening applications.
Describing own desktop usage	5	Describing Desktop Usage of the layout experience.
PC peripherals assumed	4	When describing the AR layout PC peripherals are used to navigate it.
Desktop habits acknowledged	2	Using desktop habits for the AR layout.
Preferred on a laptop	1	The layout would work better on a laptop compared to AR.
T5 Freedom versus structure		
Freedom preferred	111	Valuing unrestricted placement and sizing, including where participants added structure on top of a free space.
Space itself is valued	67	Appreciating the additional room AR provides, independent of how placement is constrained.
Restriction experienced as obstructive	57	Objecting to a constraint, particularly where visibly available space was withheld.
Overwhelmed by open space	51	Reporting that the unbounded space was too much to manage or track. The counterpart to Freedom preferred.
Restriction experienced as helpful	49	Valuing a constraint for the arrangement effort it saved or discipline it imposed.
Mental load reported	29	Explicit statements about cognitive effort in either direction, including reduced load under constraint.
Slight vertical extension wanted	4	Move the verticality of the prototype up a bit.
Vertical placement wanted	3	Want more verticality to place windows.
Reduced sense of control	2	Constraints made users feel less in control of placing a layout.
Prototype and apparatus		
Window sizes too limited	16	Wanting sizes beyond the three the prototype offered, in either direction.
Prototype fidelity commented on	5	Remarks on how well the physical prototype stood in for an AR system, positive or negative.
Fixed window height criticised	2	Objecting to the constant window height or the missing vertical dimension of the prototype.
Single window size criticized	1	Having a fixed size is seen as a negative experience.
Curved view preferred	1	Windows curving around is preferred compared to a straight line.
VR preferred over AR	1	Wants to use the layout in a VR setting instead of AR.
Concept video reactions		
Video matched expectation	21	Confirming that the behaviour shown in the concept video was what the participant had imagined.
Video diverged from expectation	15	Reporting a mismatch with the concept video, or a changed opinion after seeing it.
Source marker		
Drawing transcription	126	This highlight is used to show that a highlight is from the transcription of the drawings.

Appendix C

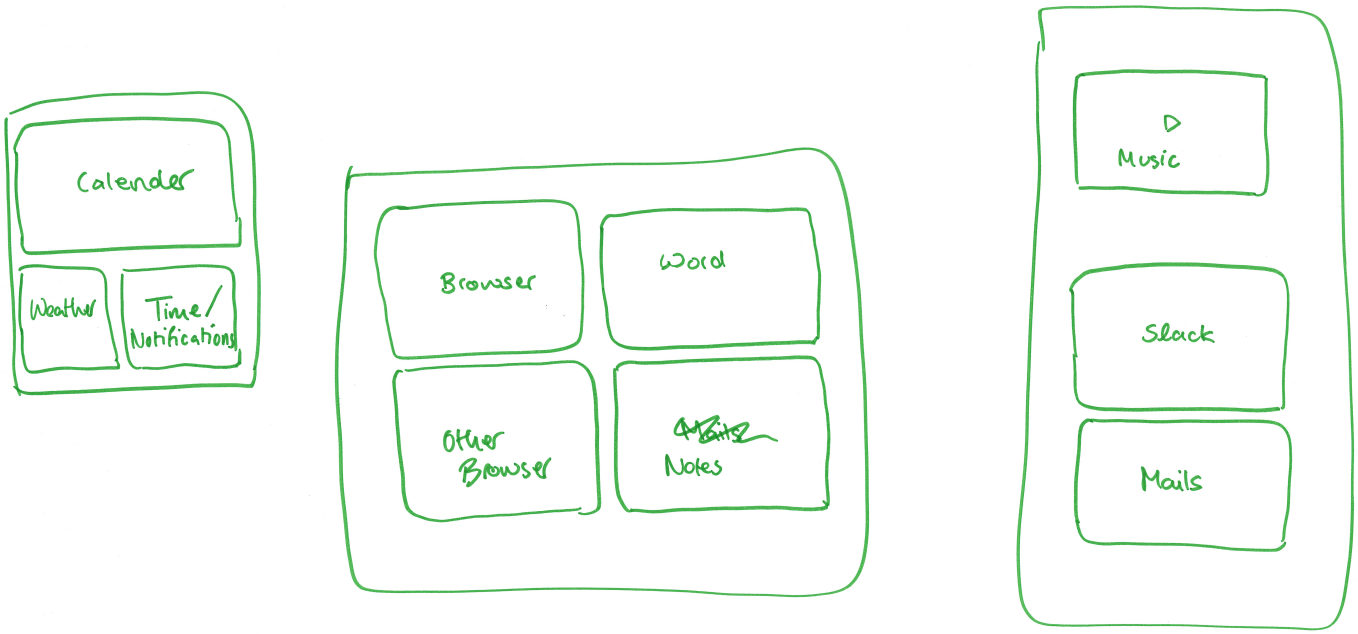
Drawings Study 1

Full size images of the drawings provided in Figure 3.11.

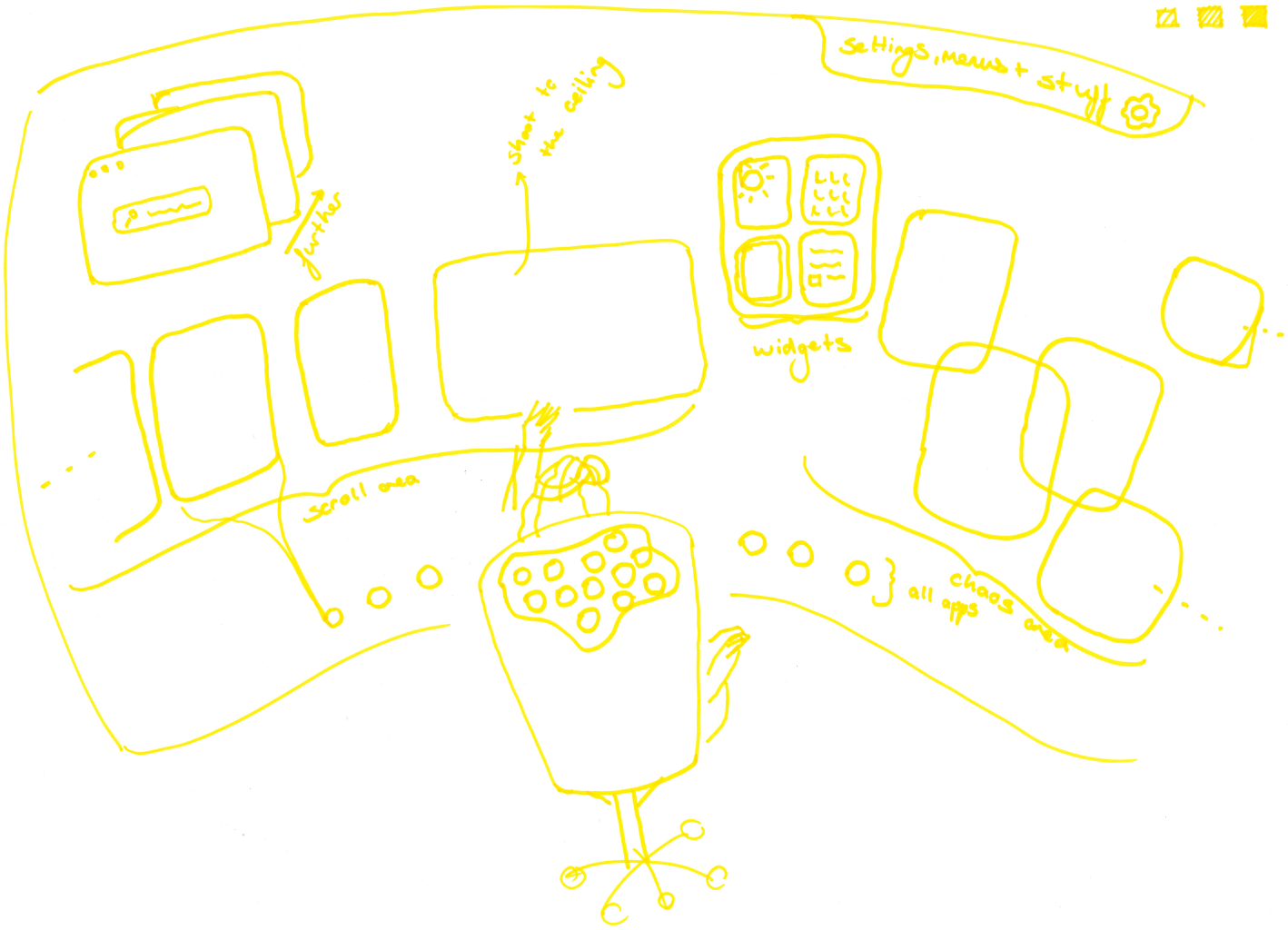
Upper-Left


low interaction auxilloy
main task area
immediate notifications and states
Interaction section messages / Notifications
Used for Communications or Notifications that are active or need immediate attention like Calls with ppl
Messages Group
Slack
main task
20°C
reference pages (datasheet / or similar
Calendar
Grouped Any Calendar events...
1
2

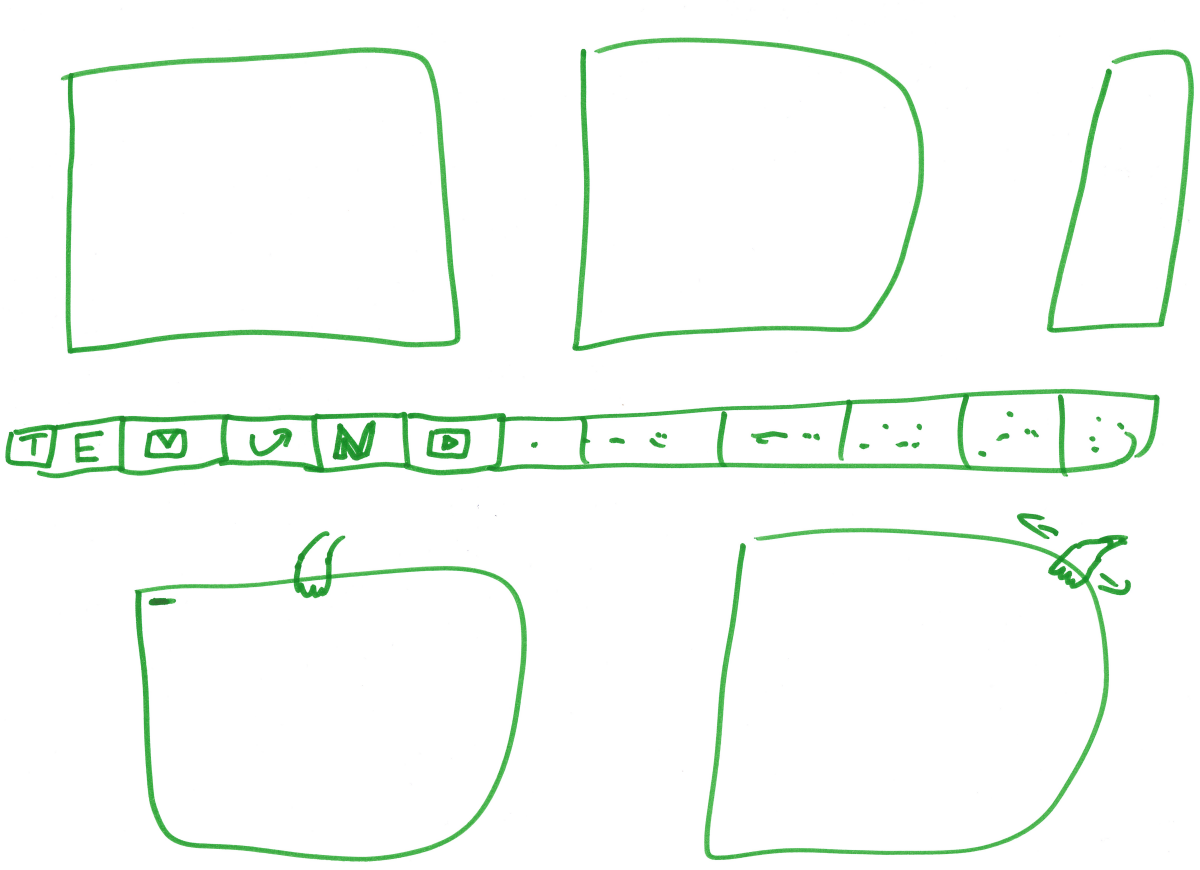
Upper-Right


Calender
Weather
Time/
Notifications
Browser
Word
Other
Browser
Notes
Music
Slack
Mails

Bottom-Left


Settings, menus + stuff
shoot to the ceiling
further
widgets
scroll area
all apps
chaos area

Bottom-Right



Appendix D

Questionnaires Study 2

In the following, we include the questionnaires used in the evaluation study (Chapter 5). Participants completed a demographics questionnaire once at the start of the session, the per-condition questionnaire after each of the three conditions, and a final questionnaire after all conditions were complete.

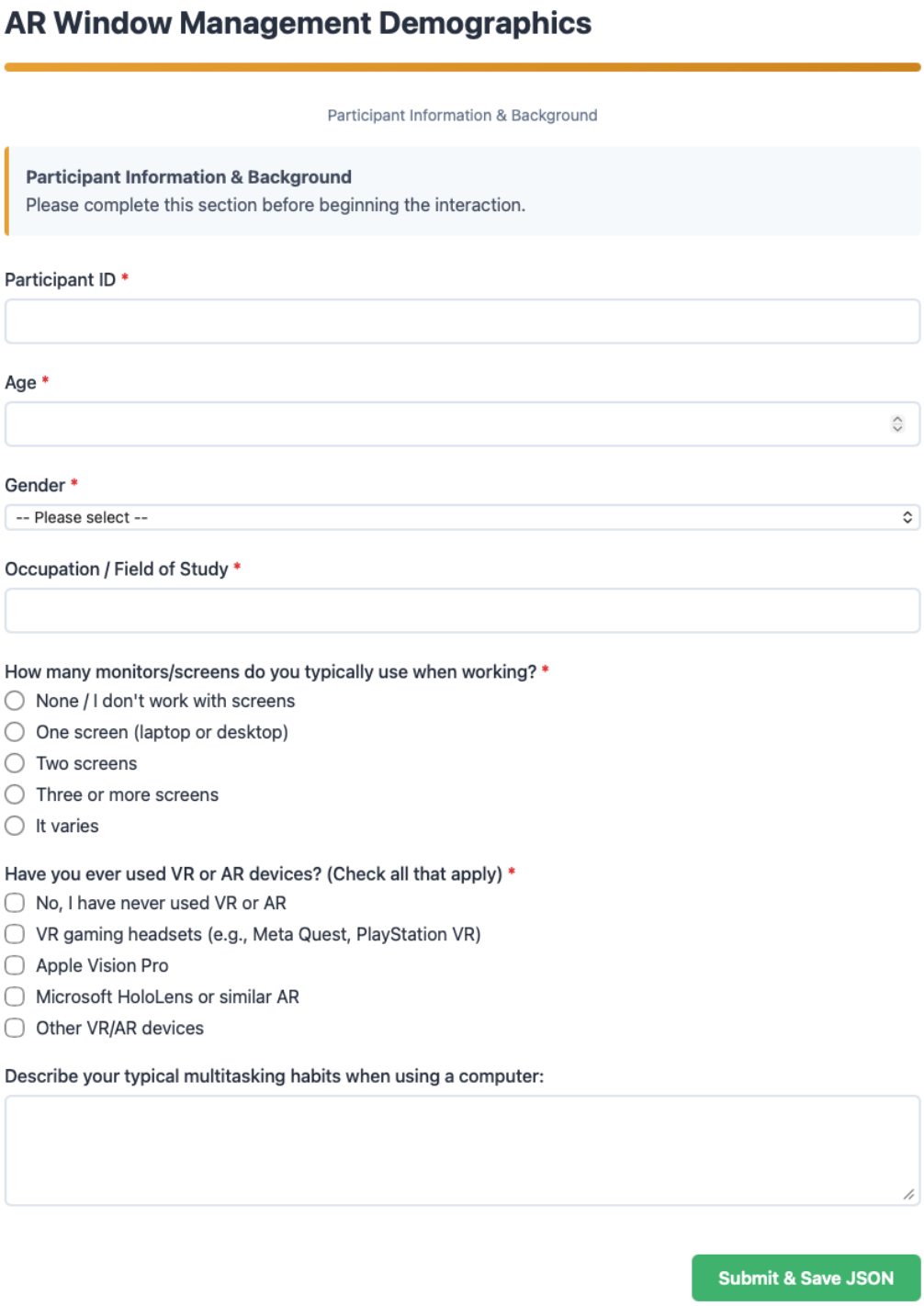


Figure D.1: Demographics questionnaire, completed once at the start of the session. The items are identical to those of the formative study, allowing the two samples to be compared.

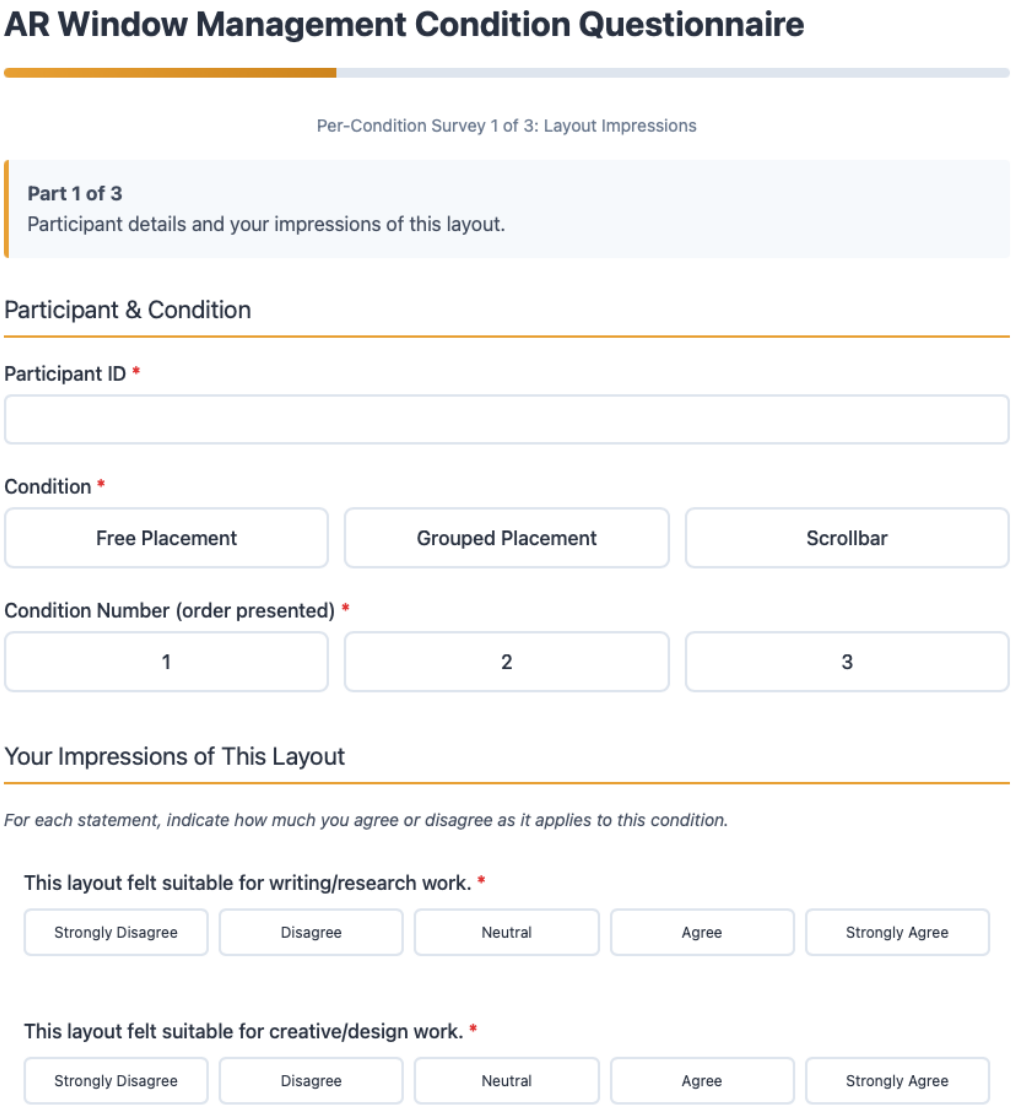


Figure D.2: Per-condition questionnaire, Part 1 of 3. Participants recorded the condition and its position in their counterbalanced order, then rated the layout on the eight items. These are the same eight items used in the formative study, which allows the ratings, in particular the sense of control, to be compared across the two studies.

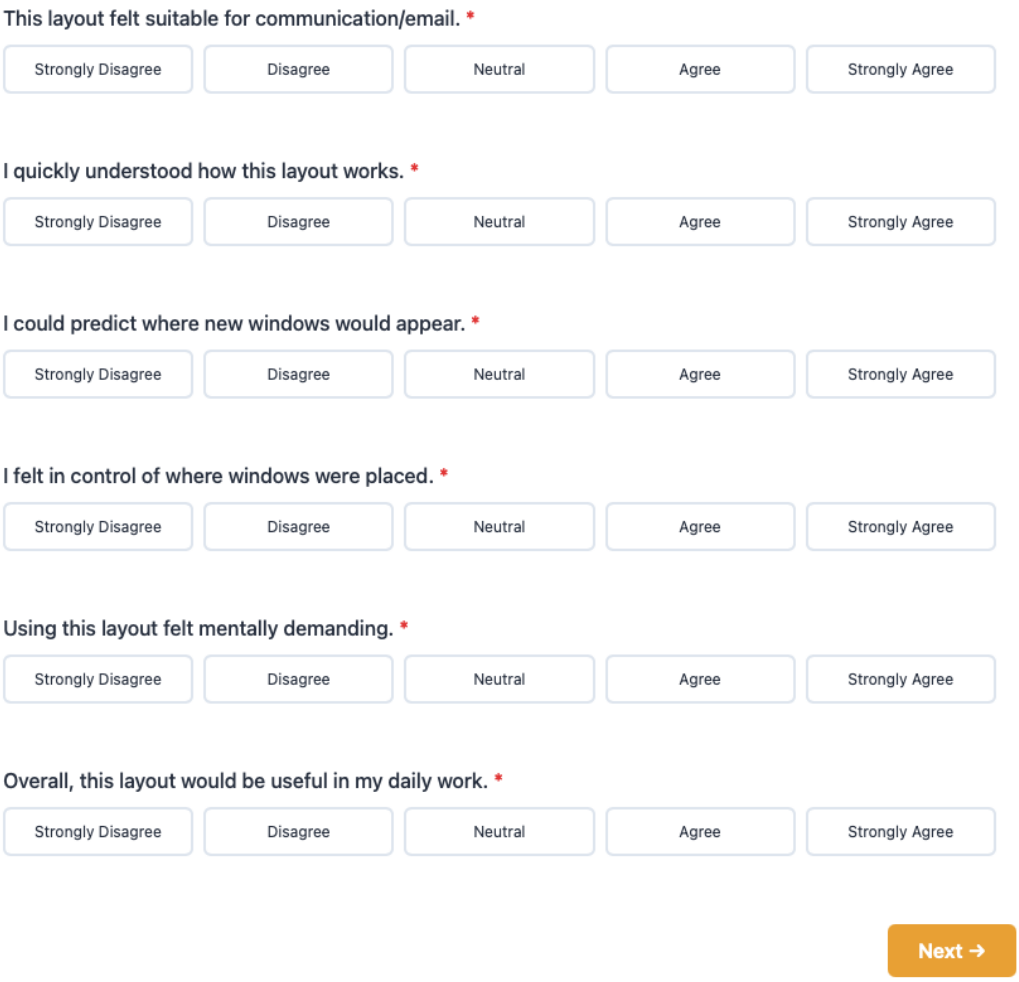


Figure D.3: The remaining six rating items of Part 1 of 3.



Figure D.4: Per-condition questionnaire, Part 2 of 3. The NASA RawTLX, rated on each of the six dimensions on a 21-point scale. Shown here are mental, physical, and temporal demand.

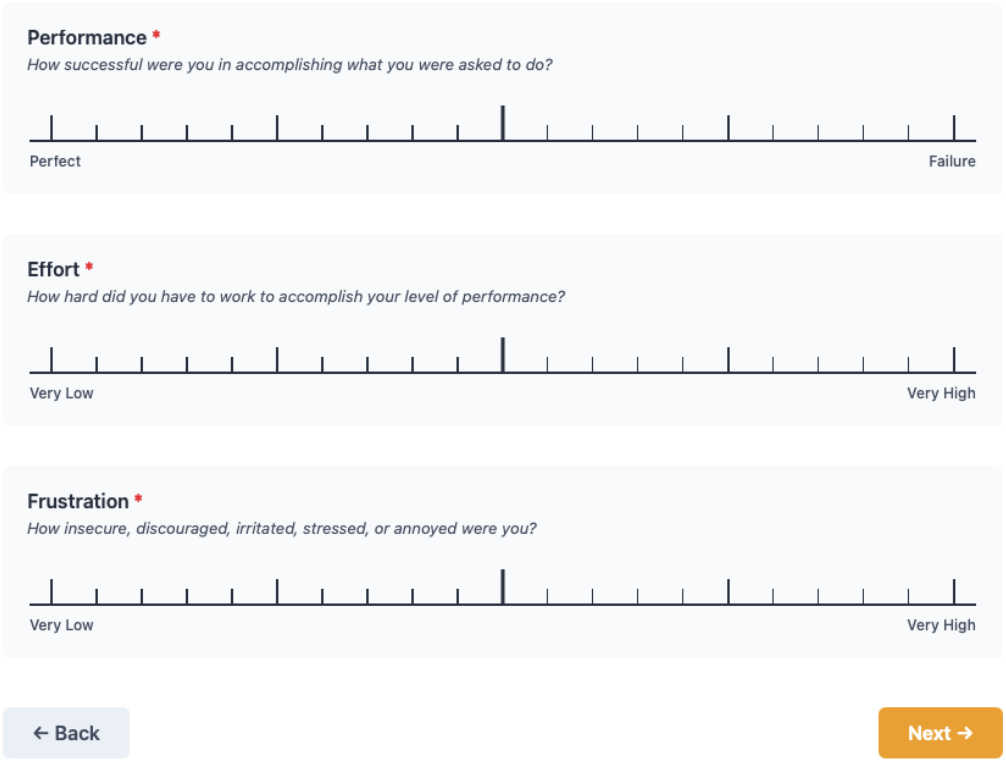


Figure D.5: The remaining three NASA RawTLX dimensions: performance, effort, and frustration.

AR Window Management Condition Questionnaire

Per-Condition Survey 3 of 3: Usability & Feedback

Part 3 of 3
System usability and open feedback.

System Usability Scale

For each statement, indicate how much you agree or disagree as it applies to this condition. Please respond based on your immediate reaction.

1. I think that I would like to use this system frequently. *

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

2. I found the system unnecessarily complex. *

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

3. I thought the system was easy to use. *

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

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

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

5. I found the various functions in this system were well integrated. *

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

6. I thought there was too much inconsistency in this system. *

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

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

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

Figure D.6: Per-condition questionnaire, Part 3 of 3. The System Usability Scale, items 1 to 7.

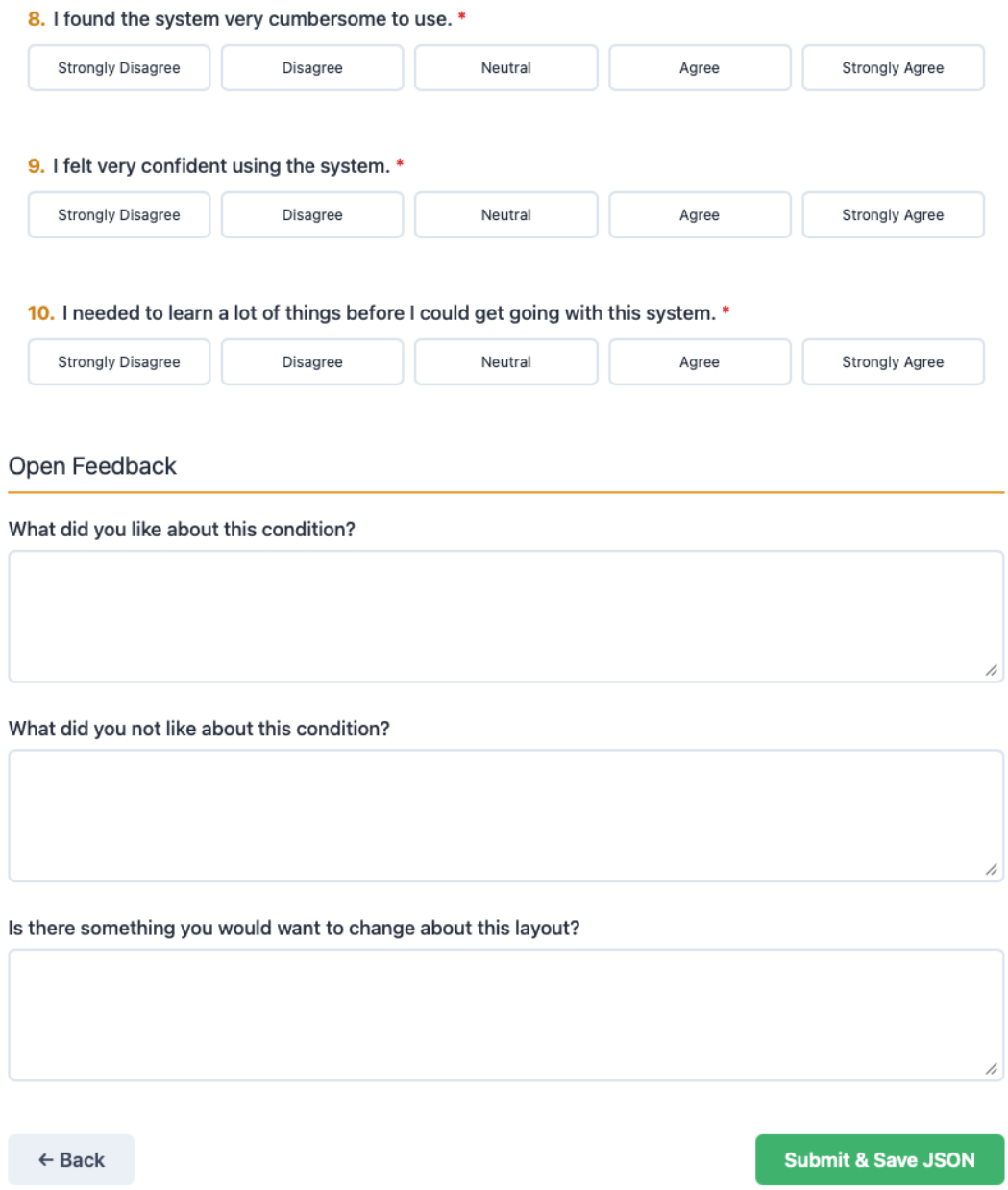


Figure D.7: Per-condition questionnaire, Part 3 of 3. The remaining System Usability Scale items and the three open questions on what participants liked, disliked, and would change. These free-text answers form the qualitative data analysed in Section 5.3.

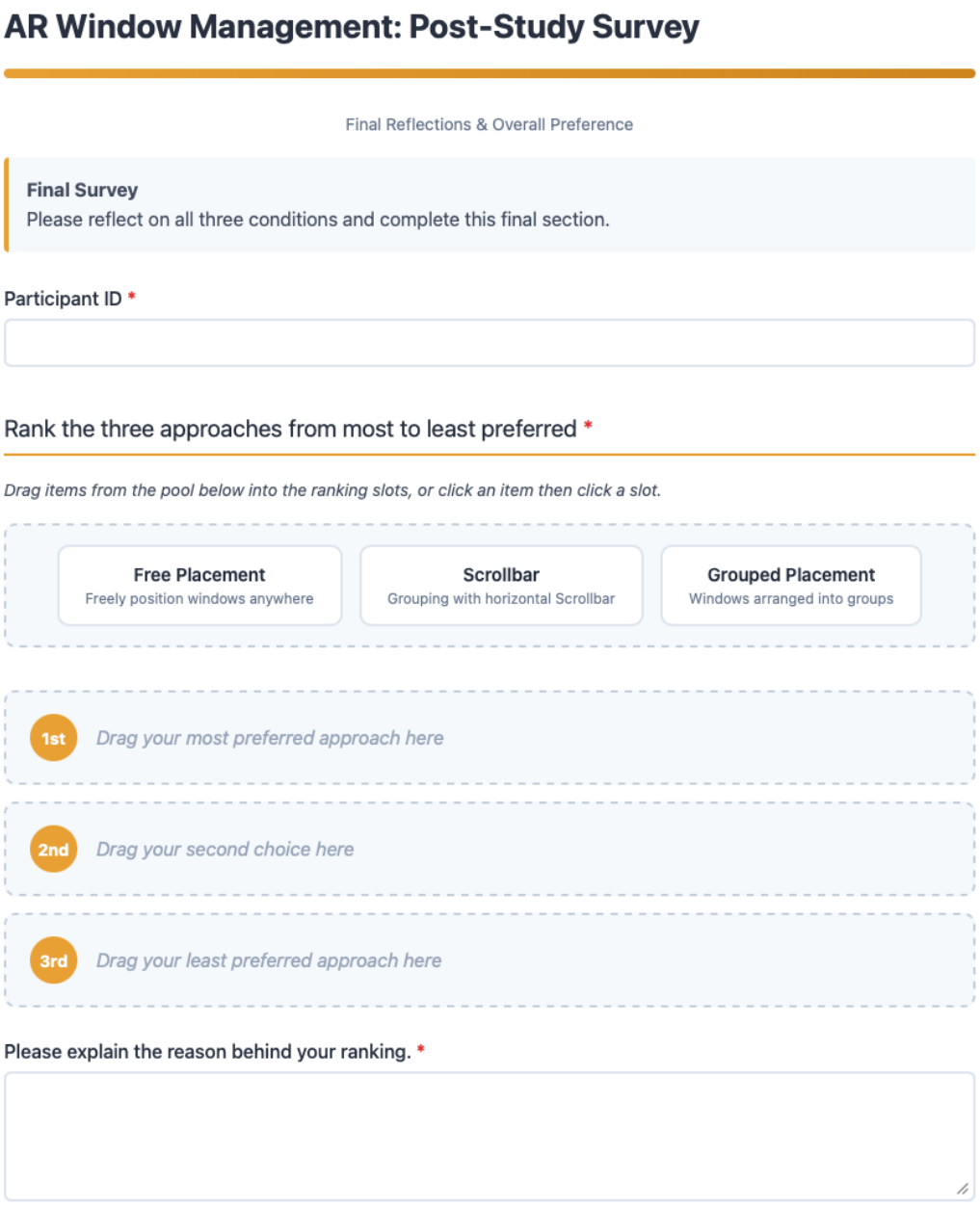


Figure D.8: Final questionnaire, completed after all three conditions. Participants ranked the conditions by overall preference and justified their ranking.

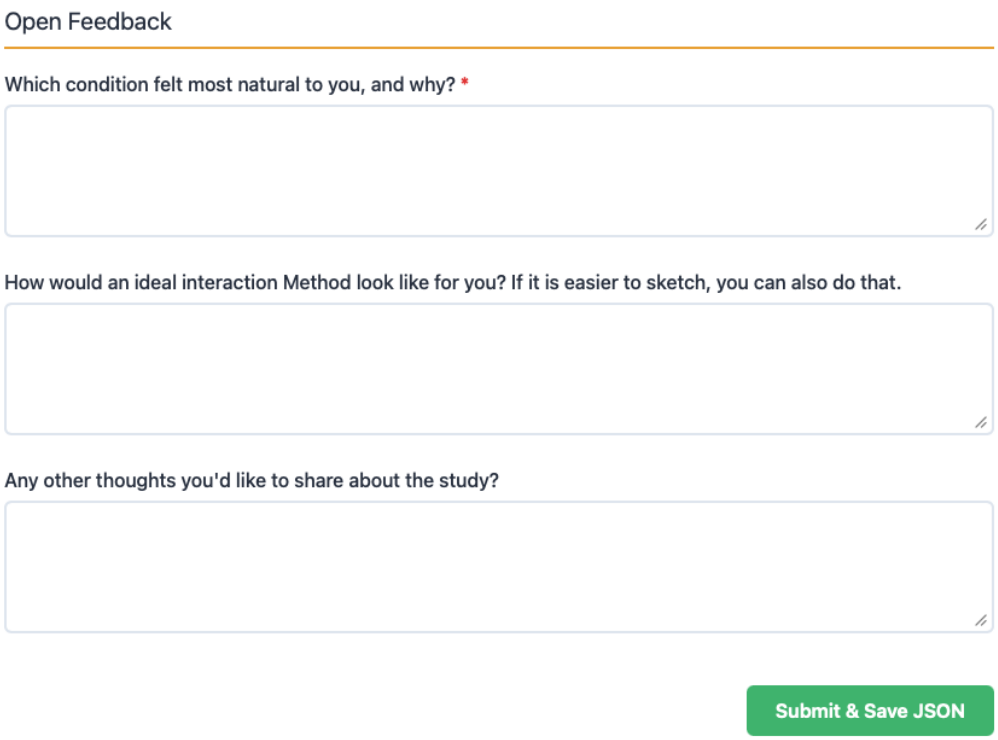


Figure D.9: The closing open questions asked which condition felt most natural, how an ideal interaction would look, and for any further thoughts on the study. Together with the per-condition open questions, these form the qualitative data analyzed in Section 5.3.

Appendix E

Codebook Study 2

The following codebook contains all codes constructed during the analysis of the free-text answers of the evaluation study, grouped under the five themes reported in Chapter 5.

Following Braun et al. [2019], themes were judged by analytic meaning rather than by frequency. Because the data are short written answers rather than interview transcripts, codes are populated more thinly than in the formative study. A segment may carry more than one code. The five themes are described in Section 5.3. Each theme is introduced by a row with an orange background. The final two groups fall outside the five themes: implementation feedback and overall reception.

- Grouping as the preferred structure
- Freedom as the baseline expectation
- The rejection of the scrolling mechanism
- The demand for flexibility inside structure
- The persistence of desktop concepts

Code	Segments	Description
Grouping as the preferred structure		
Grouping liked	20	Approving of the grouping mechanism, whether as a bare preference or with a reason given. Excludes statements that grouping felt natural because it matched an organisation participants already held, which are coded under Grouped Placement most natural.
Grouping missed in Free Placement	7	Wanting grouping back while working in Free Placement, where it was unavailable.
Grouped Placement most natural	7	Naming Grouped Placement the condition that felt most natural to use.
Color coding wanted	1	Wanting color as an additional means of marking which windows belong together.
Grouping preferred without the scrollbar	1	Preferring grouping specifically without the scrolling mechanism added on top.
Group moved as a unit	1	Valuing that a group can be repositioned as a single unit.
Grouping disliked	1	Objecting to the grouping mechanism.
Freedom as the baseline expectation		
Total control of windows	19	Reporting liking unrestricted control over placement around the whole view.
Free Placement most natural	4	Naming Free Placement the condition that felt most natural, as distinct from preferring it overall.
Free Placement preferred	4	Ranking or naming Free Placement as the preferred condition.
Snap indicator wanted in Free Placement	3	Wanting the snap preview indicator available in Free Placement as well, where it was absent.
Convenience separated from comfort	1	Separating convenience from comfort: grouping was the more efficient condition, free placement the more agreeable one.
Customization valued	1	Valuing the ability to tailor the layout to personal preference.
Placement within view appreciated	1	Appreciating that windows could be placed anywhere within view.
Surrounding space helped	1	Reporting that the surrounding space was useful rather than excessive.
All conditions felt equally natural	1	Reporting no meaningful difference in naturalness between the three conditions.
Free Placement more cumbersome	1	Finding unstructured placement more effortful.
The rejection of the scrolling mechanism		
Scrollbar rejected	10	Objecting to the scrollbar, on grounds of clutter, confusion, or redundancy.
Scrollbar ignored	4	Reporting that the mechanism was not used at all during the condition.
Vertical scrolling proposed	3	Proposing vertical scrolling within groups in addition to horizontal.
Scrollbar treated as optional extra	3	Treating the scrollbar as an optional extra that neither helps nor hinders.
Scrolling difficult to perform	2	Reporting that the scrolling interaction itself was hard to perform or understand.
Unused features cost nothing	2	Accepting the scrollbar because an unused feature costs nothing.
Scrollbar liked	2	Approving of the mechanism.
Alternative scroll interaction proposed	1	Proposing a different interaction for the scroll without rejecting the concept.
Scrolling limited to one-dimensional groups	1	Proposing that scrolling be enabled only for one dimensional groups.
Clearer scrollbar feedback wanted	1	Wanting clearer feedback that the scrollbar is active or has been triggered.
The demand for flexibility inside structure		
Automatic shifting within a group rejected	4	Objecting to existing group members shifting when a window is added.
Flexible arrangement within a group wanted	4	Wanting the arrangement within a group to be less rigid than the fixed grid.
Reordering within a group wanted	3	Wanting to reorder or reposition windows within an established group.
Stacked groups proposed	2	Proposing groups arranged as a stack rather than a grid.
Individual orientation within a group wanted	2	Wanting each window in a group to face the viewer individually rather than share the group's orientation.
Free window rotation wanted	2	Wanting manual control over window orientation, including turning a window away from the viewer.
Cumbersome to move several windows	2	Finding it effortful to reposition multiple windows.
Multi-window selection wanted	2	Wanting to move several windows at once in Free Placement.
Detaching from a group	1	Remarks on how a window is removed from a group.
Grouping by collapsing and expanding	1	Proposing that collapsing and expanding windows serve as the grouping mechanism.
Group-wide resizing wanted	1	Wanting to resize all members of a group in one action.
Flat group plane disliked	1	Objecting to a group presenting as a single flat plane.
Moving windows between groups	1	Wanting to transfer a window from one group to another directly.
Alternative grouping mechanism proposed	1	Proposing an alternative to proximity-based snapping for forming groups.
Automatic rearrangement wanted	1	Wanting the system to reflow the layout on its own.
The persistence of desktop concepts		
Continuous resizing wanted	8	Wanting continuous resizing as on the desktop, rather than the three fixed sizes.
Minimizing wanted	7	Wanting a minimize action, carried over from desktop window management.
Tab-like collapsing of groups	6	Proposing that groups collapse into tabs that can be hidden and recalled.
Desktop-like interaction described	5	Describing a desired behaviour by direct analogy to desktop interaction.
Closing windows wanted	2	Wanting a close action for windows no longer needed.
Smaller sizes wanted	1	Wanting sizes below the smallest the prototype offered.
Switchable view modes proposed	1	Proposing switchable task specific arrangements, described by analogy to virtual desktops.
Implementation feedback		
Spawn placement criticized	9	Objecting to where new windows spawned, which could place them behind existing content at a fixed depth. The most frequently raised implementation complaint.
Button design criticized	3	Objecting to the per-window controls: their size, their reliance on colour alone, and the absence of icons identifying them.
Fixed window sizes criticized	2	Objecting to the three fixed sizes the prototype offered.
Mental overload reported	2	Reporting cognitive strain, tied to the number of open windows or the difficulty of clearing space.

Code	Segments	Description
Control over spawn location wanted	1	Wanting to choose where a new window opens rather than accepting the default.
Window could not be found	1	Reporting that a window could not be found again.
Fixed size options valued	1	Valuing the fixed size options for allowing faster changes than manual resizing.
Overlap disliked	1	Objecting to windows being able to cover one another.
Arranging the full set difficult	1	Reporting that arranging the full set became difficult.
Number of open windows remarked on	1	Remarks on the number of windows the task required.
Demonstration of controls wanted	1	Wanting a better demonstration of what each control does.
Overall reception		
Prototype enjoyed	2	Spontaneous positive remarks about the system as a whole.
Search task enjoyed	2	Spontaneous positive remarks about the search task specifically.
Easy to use	1	Reporting low effort in operating the prototype.
System felt intuitive	1	Reporting that the system was intuitive to use.
Window set matched daily work	1	Confirming that the window set matched the participant's actual work.
Learning effect noticed	1	Noticing improvement across conditions over the session.

Bibliography

[1] Ronald Azuma, Yohan Baillot, Reinhold Behringer, Steven Feiner, Simon Julier, and Blair MacIntyre. Recent Advances in Augmented Reality. *IEEE Computer Graphics and Applications*, 21(6):34–47, 2001. doi.org/10.1109/38.963459.

[2] Robert Ball and Chris North. The Effects of Peripheral Vision and Physical Navigation on Large Scale Visualization. In *Proceedings of Graphics Interface 2008*, GI '08, pages 9–16, Toronto, ON, Canada, 2008. Canadian Information Processing Society.

[3] Robert Ball, Chris North, and Doug A. Bowman. Move to Improve: Promoting Physical Navigation to Increase User Performance with Large Displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, pages 191–200, New York, NY, USA, 2007. Association for Computing Machinery. doi.org/10.1145/1240624.1240656.

[4] Aaron Bangor, Philip T. Kortum, and James T. Miller. An Empirical Evaluation of the System Usability Scale. *International Journal of Human-Computer Interaction*, 24(6):574–594, 2008. doi.org/10.1080/10447310802205776.

[5] Blaine Bell, Steven Feiner, and Tobias Höllerer. View Management for Virtual and Augmented Reality. In *Proceedings of the 14th Annual ACM Symposium on User Interface Software and Technology*, UIST '01, pages 101–110, New York, NY, USA, 2001. Association for Computing Machinery. doi.org/10.1145/502348.502363.

[6] Sara A. Bly and Jarrett K. Rosenberg. A Comparison of Tiled and Overlapping Windows. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '86, pages 101–106, New York, NY, USA, 1986. Association for Computing Machinery. doi.org/10.1145/22627.22356.

[7] James V. Bradley. Complete Counterbalancing of Immediate Sequential Effects in a Latin Square Design. *Journal of the American Statistical Association*, 53(282):525–528, 1958. doi.org/10.1080/01621459.1958.10501456.

[8] Virginia Braun and Victoria Clarke. Using Thematic Analysis in Psychology. *Qualitative Research in Psychology*, 3(2):77–101, 2006. doi.org/10.1191/1478088706qp063oa.

[9] Virginia Braun and Victoria Clarke. Reflecting on Reflexive Thematic Analysis. *Qualitative Research in Sport, Exercise and Health*, 11(4):589–597, 2019. doi.org/10.1080/2159676X.2019.1628806.

[10] Virginia Braun, Victoria Clarke, Nikki Hayfield, and Gareth Terry. Thematic Analysis. In Pranee Liamputtong, editor, *Handbook of Research Methods in Health Social Sciences*, pages 843–860. Springer Singapore, Singapore, 2019. doi.org/10.1007/978-981-10-5251-4_103.

[11] John Brooke. SUS: A Quick and Dirty Usability Scale. In Patrick W. Jordan, Bruce Thomas, Bernard A. Weerdmeester, and Ian L. McClelland, editors, *Usability Evaluation in Industry*, pages 189–194. Taylor & Francis, London, UK, 1995.

[12] Ellis S. Cohen, Edward T. Smith, and Lee A. Iverson. Constraint-Based Tiled Windows. *IEEE Computer Graphics and Applications*, 6(5):35–45, 1986. doi.org/10.1109/MCG.1986.276790.

[13] Nelson Cowan. The Magical Number 4 in Short-Term Memory: A Reconsideration of Mental Storage Capacity. *Behavioral and Brain Sciences*, 24(1):87–114, 2001. doi.org/10.1017/S0140525X01003922.

[14] Mary Czerwinski, Eric Horvitz, and Susan Wilhite. A Diary Study of Task Switching and Interruptions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '04, pages 175–182, New York, NY, USA, 2004. Association for Computing Machinery. doi.org/10.1145/985692.985715.

[15] Emma E. H. Doyle, Sara E. Harrison, Stephen R. Hill, Matt Williams, Douglas Paton, and Ann Bostrom. Eliciting Mental Models of Science and Risk for Disaster Communication: A Scoping Review of Methodologies. *International Journal of Disaster Risk Reduction*, 77:103084, 2022. doi.org/10.1016/j.ijdrr.2022.103084.

[16] Barrett M. Ens, Rory Finnegan, and Pourang P. Irani. The Personal Cockpit: A Spatial Interface for Effective Task Switching on Head-Worn Displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pages 3171–3180, New York, NY, USA, 2014. Association for Computing Machinery. doi.org/10.1145/2556288.2557058.

[17] Steven Feiner, Blair MacIntyre, Marcus Haupt, and Eliot Solomon. Windows on the World: 2D Windows for 3D Augmented Reality. In *Proceedings of the*

6th Annual ACM Symposium on User Interface Software and Technology, UIST '93, pages 145–155, New York, NY, USA, 1993. Association for Computing Machinery. doi.org/10.1145/168642.168657.

[18] Victor M. González and Gloria Mark. "Constant, Constant, Multi-Tasking Craziness": Managing Multiple Working Spheres. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '04, pages 113–120, New York, NY, USA, 2004. Association for Computing Machinery. doi.org/10.1145/985692.985707.

[19] Robin S. Grenier and Dana Dudzinska-Przesmitzki. A Conceptual Model for Eliciting Mental Models Using a Composite Methodology. *Human Resource Development Review*, 14(2):163–184, 2015. doi.org/10.1177/1534484315575966.

[20] Jens Grubert, Eyal Ofek, Michel Pahud, and Per Ola Kristensson. The Office of the Future: Virtual, Portable, and Global. *IEEE Computer Graphics and Applications*, 38(6):125–133, 2018. doi.org/10.1109/MCG.2018.2875609.

[21] Sandra G. Hart. NASA-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 50(9):904–908, 2006. doi.org/10.1177/154193120605000909.

[22] Sandra G. Hart and Lowell E. Staveland. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In Peter A. Hancock and Najmedin Meshkati, editors, *Human Mental Workload*, volume 52 of *Advances in Psychology*, pages 139–183. North-Holland, Amsterdam, The Netherlands, 1988. doi.org/10.1016/S0166-4115(08)62386-9.

[23] Sebastian Hueber, Eunae Jang, and Jan Borchers. Attentive Notifications: Minimizing Distractions of Mobile Notifications through Gaze Tracking. In *Proceedings of the 25th International Conference on Mobile Human-Computer Interaction*, MobileHCI '23 Companion, New York, NY, USA, 2023. Association for Computing Machinery. doi.org/10.1145/3565066.3608695.

[24] Sebastian Hueber, Johannes Wilhelm, René Schäfer, Simon Voelker, and Jan Borchers. User-Aware Rendering: Merging the Strengths of Device- and User-Perspective Rendering in Handheld AR. *Proceedings of the ACM on Human-Computer Interaction*, 7(MHCI), 2023. doi.org/10.1145/3604278.

[25] Natalie A. Jones, Helen Ross, Timothy Lynam, Pascal Perez, and Anne Leitch. Mental Models: An Interdisciplinary Synthesis of Theory and Methods. *Ecology and Society*, 16(1), 2011. doi.org/10.5751/ES-03802-160146.

[26] Eser Kandogan and Ben Shneiderman. Elastic Windows: Evaluation of Multi-Window Operations. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, CHI '97, pages 250–257, New York, NY,

USA, 1997. Association for Computing Machinery. doi.org/10.1145/258549.258720.

[27] Ruogu Kang, Laura Dabbish, Nathaniel Fruchter, and Sara Kiesler. "My Data Just Goes Everywhere": User Mental Models of the Internet and Implications for Privacy and Security. In *Proceedings of the Eleventh USENIX Conference on Usable Privacy and Security*, SOUPS '15, pages 39–52, Berkeley, CA, USA, 2015. USENIX Association.

[28] Anne Köpsel and Nikola Bubalo. Benefiting from Legacy Bias. *Interactions*, 22 (5):44–47, 2015. doi.org/10.1145/2803169.

[29] Joseph J. LaViola. A Discussion of Cybersickness in Virtual Environments. *ACM SIGCHI Bulletin*, 32(1):47–56, 2000. doi.org/10.1145/333329.333344.

[30] Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. *Research Methods in Human-Computer Interaction*. Morgan Kaufmann, Cambridge, MA, USA, 2nd edition, 2017. ISBN 978-0-12-805390-4.

[31] Weizhou Luo, Mats Ole Ellenberg, Marc Satkowski, and Raimund Dachselt. Documents in Your Hands: Exploring Interaction Techniques for Spatial Arrangement of Augmented Reality Documents. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, CHI '25, New York, NY, USA, 2025. Association for Computing Machinery. doi.org/10.1145/3706598.3713518.

[32] Emeric Marguet, Marcel Borowski, Janus Bager Kristensen, Clemens Nylandsted Klokmose, and Niklas Elmqvist. WindowSpace: A Web-Based XR Window Manager for Interacting with 2D Windows in Immersive 3D Space. *Proceedings of the ACM on Human-Computer Interaction*, 9(8), 2025. doi.org/10.1145/3773063.

[33] Ana Maria Marhan, Mihai Ioan Micle, Camelia Popa, and Georgeta Preda. A Review of Mental Models Research in Child-Computer Interaction. *Procedia - Social and Behavioral Sciences*, 33:368–372, 2012. doi.org/10.1016/j.sbspro.2012.01.145.

[34] Paul Milgram and Fumio Kishino. A Taxonomy of Mixed Reality Visual Displays. *IEICE Transactions on Information Systems*, E77-D(12):1321–1329, 1994.

[35] George A. Miller. The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. *Psychological Review*, 63 (2):81–97, 1956. doi.org/10.1037/h0043158.

[36] Meredith Ringel Morris, Andreea Danielescu, Steven Drucker, Danyel Fisher, Bongshin Lee, m. c. schraefel, and Jacob O. Wobbrock. Reducing Legacy Bias in Gesture Elicitation Studies. *Interactions*, 21(3):40–45, 2014. doi.org/10.1145/2591689.

[37] Brad A. Myers. Window Interfaces: A Taxonomy of Window Manager User Interfaces. *IEEE Computer Graphics and Applications*, 8(5):65–84, 1988. doi.org/10.1109/38.7762.

[38] Aziz Niyazov, Barrett Ens, Kadek Ananta Satriadi, Nicolas Mellado, Loic Barthe, Tim Dwyer, and Marcos Serrano. User-Driven Constraints for Layout Optimisation in Augmented Reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, New York, NY, USA, 2023. Association for Computing Machinery. doi.org/10.1145/3544548.3580873.

[39] Donald A. Norman. Some Observations on Mental Models. In Dedre Gentner and Albert L. Stevens, editors, *Mental Models*, pages 7–14. Lawrence Erlbaum Associates, Hillsdale, NJ, USA, 1983.

[40] Donald A. Norman. *The Design of Everyday Things*. Basic Books, New York, NY, USA, 2002. ISBN 978-0-465-06710-7.

[41] Leonardo Pavanatto, Feiyu Lu, Chris North, and Doug A. Bowman. Multiple Monitors or Single Canvas? Evaluating Window Management and Layout Strategies on Virtual Displays. *IEEE Transactions on Visualization and Computer Graphics*, 31(3):1713–1730, 2025a. doi.org/10.1109/TVCG.2024.3368930.

[42] Leonardo Pavanatto, Jens Grubert, and Doug A. Bowman. Spatial Bar: Exploring Window Switching Techniques for Large Virtual Displays. In *2025 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, VR '25, pages 186–194, Los Alamitos, CA, USA, 2025b. IEEE Computer Society. doi.org/10.1109/VR59515.2025.00043.

[43] Ken Pfeuffer, Benedikt Mayer, Diako Mardanbegi, and Hans Gellersen. Gaze + Pinch Interaction in Virtual Reality. In *Proceedings of the 5th Symposium on Spatial User Interaction*, SUI '17, pages 99–108, New York, NY, USA, 2017. Association for Computing Machinery. doi.org/10.1145/3131277.3132180.

[44] Marc Rettig. Prototyping for Tiny Fingers. *Communications of the ACM*, 37(4): 21–27, 1994. doi.org/10.1145/175276.175288.

[45] Jeff Sauro and James R. Lewis. *Quantifying the User Experience: Practical Statistics for User Research*. Morgan Kaufmann, Cambridge, MA, USA, 2nd edition, 2016. ISBN 978-0-12-802308-2.

[46] Carolyn Snyder. *Paper Prototyping: The Fast and Easy Way to Design and Refine User Interfaces*. Morgan Kaufmann, San Francisco, CA, USA, 2003. ISBN 978-1-55860-870-2.

[47] Nancy Staggers and Anthony F. Norcio. Mental Models: Concepts for Human-Computer Interaction Research. *International Journal of Man-Machine Studies*, 38(4):587–605, 1993. doi.org/10.1006/imms.1993.1028.

[48] Ivan E. Sutherland. A Head-Mounted Three Dimensional Display. In *Proceedings of the December 9–11, 1968, Fall Joint Computer Conference, Part I*, AFIPS '68 (Fall, part I), pages 757–764, New York, NY, USA, 1968. Association for Computing Machinery. doi.org/10.1145/1476589.1476686.

[49] Susanne Tak, Andy Cockburn, Keith Humm, David Ahlström, Carl Gutwin, and Joey Scarr. Improving Window Switching Interfaces. In *Human-Computer Interaction – INTERACT 2009*, INTERACT '09, pages 187–200, Berlin, Heidelberg, Germany, 2009. Springer. doi.org/10.1007/978-3-642-03658-3_25.

[50] Uta Wagner, Mathias N. Lystbæk, Pavel Manakhov, Jens Emil Sloth Grønbæk, Ken Pfeuffer, and Hans Gellersen. A Fitts' Law Study of Gaze-Hand Alignment for Selection in 3D User Interfaces. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, New York, NY, USA, 2023. Association for Computing Machinery. doi.org/10.1145/3544548.3581423.

[51] Yushi Wei, Rongkai Shi, Difeng Yu, Yihong Wang, Yue Li, Lingyun Yu, and Hai-Ning Liang. Predicting Gaze-Based Target Selection in Augmented Reality Headsets Based on Eye and Head Endpoint Distributions. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, New York, NY, USA, 2023. Association for Computing Machinery. doi.org/10.1145/3544548.3581042.

[52] Evan J. Williams. Experimental Designs Balanced for the Estimation of Residual Effects of Treatments. *Australian Journal of Scientific Research, Series A: Physical Sciences*, 2(2):149–168, 1949. doi.org/10.1071/CH9490149.

Index

AR ..... *see* Augmented Reality
Augmented Reality 2

Bounded Freedom 51

Carousel Metaphor 41
Condition
- Elastic Horizontal 22
- Elastic Scroll ..... *see* Elastic Horizontal
- Fixed Horizontal 21
- Free Placement (Study 1) 21
- Free Placement (Study 2) 57
- Grouped Placement 57
- Infinite Scroll ..... *see* Fixed Horizontal
- Scrollbar 57

Cybersickness 61

Degrees of Constraint 21

Extended Reality 4

Head-Worn Display 1
HWD ..... *see* Head-Worn Display
Hypothesis 68

Knowledge Worker 10

Legacy Bias 17

Mental Model 3

NASA TLX 68

RawTLX . . . . . *see* NASA TLX
Reflexive Thematic Analysis 16

Spatial Memory 10
Supplementary Material 7
SUS . . . . . *see* System Usability Scale
System
- Elastic Windows 10
- Personal Cockpit 11
- Spatial Bar 12
- SpatialWindowPlacer 62
System Usability Scale 68
System-Suggested Grouping 99

Theme
- T1 Importance 35
- T2 Window Organization 38
- T3 Bringing Windows Into View 41
- T4 Desktop Paradigms 45
- T5 Freedom Versus Structure 46
Theme (Evaluation Study) 80–87

Virtual Reality 2
VR . . . . . *see* Virtual Reality

Window Management 3–5
- Overlapping 10
- Tiled 9
- Tiling 4
Working Spheres 10

XR . . . . . *see* Extended Reality

Typeset August 12, 2026